REPORT RESUMES

ED 003 046 24

LIKELIHOOD RATIO TESTS OF HYPOTHESES ON MULTIVARIATE POPULATIONS, VOLUME 1, DISTRIBUTION THEORY--STATISTICAL MODELS FOR THE EVALUATION AND INTERPRETATION OF EDUCATIONAL CRITERIA, PART 4.

BY- SAW, J.G.

VIRGINIA POLYTECHNIC INST., BLACKSBURG

REPORT NUMBER CRP-1132 PT. 4 PUB DATE 64

EDRS PRICE MF-$0.36 HC-$8.76 219P.

DESCRIPTORS- MATHEMATICS, MATHEMATICAL MODELS, *STATISTICAL ANALYSIS, *HYPOTHESES TESTING, BLACKSBURG, VIRGINIA

THIS VOLUME DEALS WITH THE BIVARIATE NORMAL DISTRIBUTION. THE AUTHOR MAKES A DISTINCTION BETWEEN DISTRIBUTION AND DENSITY FROM WHICH HE DEVELOPS THE CONSEQUENCES OF THIS DISTINCTION FOR HYPOTHESIS TESTING. OTHER ENTRIES IN THIS SERIES ARE ED 003 044 AND ED 003 045. (JK)

ED003046

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

STATISTICAL MODELS FOR THE EVALUATION AND INTERPRETATION OF EDUCATIONAL CRITERIA

Cooperative Research Project Number 1132

by

J. G. Saw

Virginia Polytechnic Institute
Blackshurg, Virginia
1964

PART 4. LIKELIHOOD RATIO TESTS OF HYPOTHESES ON MULTIVARIATE POPULATIONS

VOLUME I. DISTRIBUTION THEORY

The presentation of these notes was supported through the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

PRECEDING PAGE MISSING

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

ACKNOWLEDGMENTS

The writer wishes to acknowledge the financial support of the U. S. Department of Health, Education, and Welfare which has greatly accelerated the preperation of these notes.

It is impossible to work in the field of multivariate theory and analysis without being aware of the tremendous contributions of several authors; in particular, those of T. W. Anderson and the late S. N. Roy. This writer wishes to acknowledge that he has been influenced by their publications.

Finally, I wish to pay tribute to the excellent technical typing of Mrs. Dottie Gladstone and of Mrs. Marty Weihe and to their ability to deal with what was a difficult manuscript.

INDEX

Chapter 1 INTRODUCTION AND NOTATION

1.1 Introduction 12

1.2 Notation 21

Chapter 2 THE MULTIVARIATE NORMAL DENSITY

2.1 Introduction 24

2.2 The multivariate normal density. 24

2.3 The marginal densities. 27

2.4 The conditional densities. 31

2.5 The canonical correlation coefficient. 33

2.6 The multiple correlation coefficient. 35

2.7 The multivariate normal distribution. 37

2.8 The distribution of a matrix of normal variables. 41

Chapter 3 THE WISHART DISTRIBUTION

3.1 Introduction 44

3.2 Definition of the Wishart distribution. 44

3.3 The generalized idempotent quadratic form. 46

3.4 The Wishart density (integer degrees of freedom). 48

3.5 The distribution of the trace of a Wishart variate. 57

3.6 Functions of the Wishart variate having a Wishart distribution. 58

3.7 The distribution of the determinant of a Wishart variate. 65

3.8 The Wishart density (non-integer degrees of freedom). 67

Chapter 4 THE U-STATISTIC; HOTELLING'S T^2

4.1 Introduction 76
4.2 Definition of the U-statistic; Moments of central U. 76
4.3 The noncentral U-variate. 81
4.4 The relationship between $U'_{m:\nu_1:\nu_2}$ and $U'_{\nu_1:m:\nu_1+\nu_2-m}$. 82
4.5 Hotelling's T^2. 87
4.6 Approximations to the distribution of the central U-statistic. 90
4.7 An inequality on the cumulative integral for the noncentral U-variate 97
4.8 The "step-down" property of $U'_{m:\nu_1:\nu_2}$. 102

Chapter 5 MOMENT GENERATING FUNCTIONS

5.1 Introduction 107
5.2 The moment generating function of a normally distributed vector. 107
5.3 The moment generating function of a noncentrally distributed Wishart variate. 109
5.4 The moment generating function of the Wishart distribution with possibly non-integer degrees of freedom. 111

Chapter 6 LATENT ROOTS AND VECTORS ASSOCIATED WITH WISHART VARIATES

6.1 Introduction 113

6.2 The distribution of latent roots 114

Chapter 7 THE GENERAL LINEAR HYPOTHESIS

7.1 Introduction 123

7.2 Likelihood estimation in the unrestricted parameter space (Ω). 124

7.3 Distribution of the unrestricted likelihood estimators. 128

7.4 Likelihood estimation in a restricted parameter space (ω). Estimable matrices. 131

7.5 Distribution of the restricted likelihood estimators. 134

7.6 The likelihood ratio test function and its distribution. 138

7.7 Hotelling's statistic in tests of linear hypothesis. 144

7.8 An alternative system of null hypotheses. 146

Chapter 8 SOME PROPERTIES OF THE LIKELIHOOD ESTIMATES

8.1 Introduction 151

8.2 The geometry of likelihood estimation. 151

8.3 The geometry of likelihood ratio tests. 155

8.4 The accuracy of a vector of normally distributed estimators. 159

Chapter 9 TEST OF HYPOTHESIS CONCERNING THE FORM OF THE DISPERSION MATRIX

9.1 Introduction 165

9.2 To test that the dispersion matrix is equal to a fully specified; H_o: $V = V_o$. 166

9.3 To test that a dispersion matrix is proportional to a specified matrix; H_o: $V \alpha V_o$. 167

9.4 To test that a number of dispersion matrices are equal (test of homoscedasticity). 171

9.5 To test for independence between two groups of observations. 175

Exercises 179

Appendix A SOME RESULTS IN THE THEORY OF MATRICES AND DETERMINANTS

A.1 Introduction 199

A.2 The clockwise rule. 199

A.3 The inverse of a partitioned matrix. 201

A.4 The vector and matrix differential operator. 202

A.5 The Kronecker product of matrices. 205

Appendix B THE JACOBIAN OF A TRANSFORMATION USED IN CHAPTER 6

B.1 Introduction 207

B.2 The evaluation. 207

B.3 The technique of linear equivalence. 211

Bibliography 213

PREFACE

In preparing these notes it was hoped they would meet the needs of the advanced graduate student and those who have a sound background in the theory of tests of linear hypothesis in the univariate case (to the level of say, Graybill's *An Introduction to Linear Statistical Models, Volume I*). Some of the topics herein will certainly lie outside a five-credit course in multivariate distribution theory and proceedure. Other topics which some would perhaps have included will not be found in these notes; for example, elements of factor analysis (which are discussed and applied by Rolf Bargmann under Part III of this contract) are omitted.

Examples and illustrations were thought to be unnecessary and even undesirable in this first volume since we will repeatedly rely on the univariate theory for fortification and direction. Our object should be to generalize the univariate theory to the case where a set of vector observations are available rather than a set of scalar observations. Thus we can build up a complete theory in the multivariate case using as template all that has been learned in the univariate situation. The spirit of this volume is then a study of distribution theory for its own sake. Those who would wish to keep a foot on the ground may choose to read Volume II collaterally.

Some topics which are touched upon are the product of very recent research. A significant paper by S. Das Gupta, T. W. Anderson, and G. S. Mudholkar which appears in the March 1964 volume of Annals of Mathematical Statistics is adapted in a simple case in Section 4. This and other recent developments have been included in the hope of stimulating the reader to possible research in the field. There is yet much investigation to be done; for example, in connexion with the power of likelihood ratio tests of hypothesis.

The distribution theory herein discussed will be relevant to likelihood ratio tests or data drawn from an underlying normal population. Most of the data which reaches us can indeed be considered to be drawn from an underlying population which has a normal density (or an approximately normal density) so that our considerations will have great utility.

CHAPTER 1

INTRODUCTION AND NOTATION

1.1 Introduction

Early in his reading of statistical methods and theory, the reader will meet the normal bivariate density; in fact, it is assumed that the reader of these notes has a familiarity with this surface and its properties. The standard form for this density:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp - \frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) \cdot \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}$$

(when $\sigma_x > 0$; $\sigma_y > 0$ and $|\rho| < 1.$) is described verbally as "the normal bivariate density" on variables (x,y) with means μ_x, μ_y respectively; standard deviations σ_x, σ_y respectively and a coefficient of correlation of ρ. The cumulants of this density are known to be

$$k_{10}(x,y) = \mu_x \quad ; \quad k_{01}(x,y) = \mu_y$$

$$k_{20}(x,y) = \sigma_x^2 \quad ; \quad k_{02}(x,y) = \sigma_y^2$$

$$k_{11}(x,y) = \rho\sigma_x\sigma_y$$

$$k_{r,s}(x,y) = 0 \quad r+s>2.$$

Given a series of observations $\{(x_i,y_i)\}_1^n$ from a population whose density is the f(x,y) given above, the reader is probably familiar with the methods and theory appropriate for test of such hypotheses as:

(a) $\mu_x = \mu_y$

(b) $a\mu_x + b\mu_y + c = 0$ (for given a,b,c)

(c) $\rho = 0$

(d) $\rho = \rho_0$ (for given ρ_0)

against a variety of alternatives. It may perhaps surprise the student that he has not met a description of a method for testing the hypothesis

(e) $\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \begin{bmatrix} {}_0\mu_x \\ {}_0\mu_y \end{bmatrix}$ for given ${}_o\mu_x$ and ${}_o\mu_y$. σ_x, σ_y and ρ unknown.

It is not difficult to establish the likelihood ratio test for this hypothesis against the general alternative; it is in fact the $\frac{n}{2}$th power of

$$\frac{\begin{vmatrix} \sum\limits_i (x_i-\bar{x})^2 & \sum (x_i-\bar{x})(y_i-\bar{y}) \\ \sum (x_i-\bar{x})(y_i-\bar{y}) & \sum (y_i-\bar{y})^2 \end{vmatrix}}{\begin{vmatrix} \sum (x_i-\bar{x})^2 + n(\bar{x}-{}_o\mu_x)^2 & \sum (x_i-\bar{x})(y_i-\bar{y}) + n(\bar{x}-{}_o\mu_x)(\bar{y}-{}_o\mu_y) \\ \sum (x_i-\bar{x})(y_i-\bar{y}) + n(\bar{x}-{}_o\mu_x)(\bar{y}-{}_o\mu_y) & \sum (y_i-\bar{y})^2 + n(\bar{y}-{}_o\mu_y)^2 \end{vmatrix}}$$

This is not a complicated expression either in its statement or respect of its computation and yet clearly some new results in the distribution theory are likely to be needed to cope with criteria which are expressed as determinants of an array of random variables.

Of course the bivariate distribution is a special case of the m-variate distribution. In a wider domain of application we will be dealing with a vector

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

of random variables; that is, $x_1,x_2,\ldots,x_m$ have a joint density and $\mathcal{E}(x_i) = \mu_i$; $\mathrm{Var}(x_i) = v_{ii}$; $\mathrm{Cov}(x_i,x_j) = v_{ij}$. The multivariate counterparts of hypotheses (a)-(e) become

(a)' $\mu_1 = \mu_2 = \cdots = \mu_m$.

(b)' $a_1\mu_1+a_2\mu_2+\cdots+a_m\mu_m+c=0$.

(c)' $v_{ij}=0$ some given i and j

(d)' $v_{ij}={}_ov_{ij}$ some given i and j

(e)' $\{\mu_i={}_o\mu_i\}_{i=1}^m$ with $\{{}_o\mu_i\}_1^m$ given

and these counterparts may be further extended:-

(b)"
$$\begin{cases} a_{11}\mu_1+a_{12}\mu_2+\cdots+a_{1m}\mu_m+c_1=0 \\ a_{21}\mu_1+a_{22}\mu_2+\cdots+a_{2m}\mu_m+c_2=0 \\ \vdots \\ a_{p1}\mu_1+a_{p2}\mu_2+\cdots+a_{pm}\mu_m+c_p=0 \end{cases}$$

(c)" $v_{ij}=0$ some given set of i and j to give just two immediately apparent extensions.

Actually the methods of dealing with these extensions are surprisingly easy for the most part. The reader may take heart to learn for example that every test of the general linear hypothesis in the univariate case (normal density assumed) has its immediate counterpart in the m-variate case. For example:-

Problem I. Univariate classification by groups.

Given $\{x_{ij}\}_{i=1;j=1}^{k\;;n_i}$, a set of independently distributed normal variates such that

$$\mathcal{E}(x_{ij}) = \beta_i \qquad j=1,\dots,n_i \;;\; i=1,\dots,k$$

$$\mathrm{Var}(x_{ij}) = \sigma^2 \qquad \text{all } i \text{ and } j;\ \sigma^2 \text{ unknown.}$$

Required to test H_0: $\beta_1=\beta_2=.\ .\ .=\beta_k$

H_1: general alternative.

Method

Construct

$$S_h = \sum_{i=1}^{k}(\bar{x}_{i.}-\bar{x}_{..})^2 n_i$$

$$S_e = \sum_{i=1}^{k}\ \sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2$$

where $\bar{x}_{i.} = \frac{1}{n_i}\sum_j x_{ij}$; $\bar{x}_{..} = \frac{1}{N}\sum_{i=1}^{k}\ \sum_{j=1}^{n_i} x_{ij}$

with $N = \sum_{i=1}^{k} n_i$

Refer $\dfrac{S_h/k-1}{S_e/N-k}$ to $F_{k-1:N-k}$ (Fisher's variance-ratio).

Problem II. Multivariate classification by groups.

Given $\{\underline{x}_{ij}\}_{i=1;j=1}^{k\;\;n_i}$ a set of independently distributed normal column vectors of dimension m. (The vectors have independent distributions; not the elements within the vector.)

$$\mathcal{E}(\underline{x}_{ij}) = \underline{\beta}_i \qquad j=1\ldots n_i; i-1\ldots k$$

Required to test H_0: $\underline{\beta}_1=\underline{\beta}_2=\ldots\underline{\beta}_k$

H_1: general alternative.

Method

Construct $\bar{\underline{x}}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} \underline{x}_{ij}; \bar{\underline{x}}=\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\underline{x}_{ij}$

$$S_H = \sum_{i=1}^{k}(\bar{\underline{x}}_{i.}-\bar{\underline{x}}_{..})(\bar{\underline{x}}_{i.}-\bar{\underline{x}}_{..})'n_i$$

$$S_E = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\underline{x}_{ij}-\bar{\underline{x}}_{i.})(\underline{x}_{ij}-\bar{\underline{x}}_{i.})'$$

We will reveal the test function as $|S_E|/|S_E+S_H|$ but leave the distribution of this test function undisclosed.

Notice then the similarity between the univariate method and the multivariate method. The vector means $\bar{\underline{x}}_{i.}$ and $\bar{\underline{x}}$ are m-dimensional vectors; the α-th element of $\bar{\underline{x}}$ is the arithmetic mean of the α-th elements in the set of $\underline{x}_{ij}$. The squares in the univariate case become the product of a column vector by its transpose, thus

$$(\bar{x}_{i.}-\bar{x}_{..})^2 \rightarrow (\bar{\underline{x}}_i-\bar{\underline{x}}_{..})(\bar{\underline{x}}_i-\bar{\underline{x}}_{..})'$$

so that S_H and S_E are both mxm (symmetric) matrices.

Although the simplest of designs has been chosen by way of illustration, the duality between the univariate and multivariate test of linear hypothesis extends to all designs, cross-classification; latin squares; Youden squares and so forth.

The matrix S_H of problem II must include the sum of squares S_h of problem I as the special case m=1. An immediate consequence then is that the diagonal elements of S_H are χ^2-variates. It may be of help to regard the univariate theory as a "special case" of the multivariate theory; we have a constant check on our work in m-dimensions if we set m=1 when we must have a (presumably well known) result in the univariate situation. We now set out in tabular form some definitions in the multivariate theory and compare them with the corresponding definitions in the univariate theory.

Multivariate case $m > 1$.	Univariate case $m = 1$.
$\underline{x}$: an m element column vector	x : a scalar
$\underline{x}$ has a multivariate normal density (see Chapter 2.)	x has a univariate normal density
$\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_\nu$ are independent random vector variates	$x_1, x_2, \ldots, x_\nu$ are independent random scalar variates.
$\sum_{j=1}^{\nu} \underline{x}_j \underline{x}_j'$ has a Wishart distribution (see Chapter 3).	$\sum_{j=1}^{\nu} x_j^2$ has a χ^2-distribution

$\dfrac{\left\lvert \sum_{j=1}^{\nu_2} \underline{x}_j \underline{x}_j' \right\rvert}{\left\lvert \sum_{j=1}^{\nu_2} \underline{x}_j \underline{x}_j' + \sum_{j=1}^{\nu_1} \underline{y}_j \underline{y}_j \right\rvert}$ has a U- distribution. (See Chapter 4.)	$\dfrac{\sum_{j=1}^{\nu_2} x_j^2}{\sum_{j=1}^{\nu_2} x_j^2 + \sum_{j=1}^{\nu_1} y_j^2}$ has a Beta distribution.
$\overline{\underline{x}}$ is distributed independently of $\sum_{j=1}^{\nu} (\underline{x}_j - \overline{\underline{x}})(\underline{x}_j - \overline{\underline{x}})'$	$\overline{x}$ is independent of $\sum_{j=1}^{\nu} (x_j - \overline{x})^2$
$\frac{(n-1)n}{\text{[redacted]}} \overline{\underline{x}}' \left\langle \sum (\underline{x}_j - \overline{\underline{x}})(\underline{x}_j - \overline{\underline{x}})' \right\rangle^{-1} \overline{\underline{x}}$ has a T^2-distribution (See Chapter 4.)	$\dfrac{n\overline{x}^2}{\frac{1}{n-1} \sum (x_j - \overline{x})^2}$ has a Student's t^2-distribution.

From the definitions of the Wishart distribution; the U-distribution and the T^2-distribution, it is apparent that we will expect: the χ^2-distribution is a special case of the Wishart distribution; the Beta-distribution is a special case of the U-distribution; the t^2-distribution is a special case of the T^2-distribution, and this is indeed the case provided that we define the m-variate normal density in a suitable manner.

It is not immediately obvious how to extend or generalize the univariate density. We may choose to define a density on $\underline{x}$ such that the marginal distribution on every element of $\underline{x}$ has a univariate normal density. Clearly if we set m=1, we now have the special case of the univariate normal density. However, this is not too satisfactory for if $f(x_1,x_2,\ldots,x_m)$ satisfy the requirements of this paragraph, then so also does $g(x_1,x_2,\ldots,x_m) = f(x_1,x_2,\ldots,x_m) - a(x_1,x_2,\ldots,x_m)$ where $a(x_1,x_2,\ldots,x_m)$ satisfies

$$a(x_1,x_2,\ldots,x_m) \leq f(x_1,x_2,\ldots,x_m) \text{ all } \underline{x}$$

and where the integral over all variables but (say) the j-th of $a(x_1,x_2,\ldots,x_m)$ is zero for each j. Thus we may produce an essentially infinite variety of types of surfaces for the density each satisfying the requirement that the marginal densities are normal.

We choose to be a little more restrictive in our definition of the multivariate surface and demand that the distribution of every linear conbination of the elements of $\underline{x}$ has a non-singular (i.e. standard deviation strictly positive) univariate normal density. Now if $\{u_j\}_{j=1}^m$ are a set of independent normal variates then the set of $\{y_j\}_{j=1}^m$ defined by

$$\underline{y} = G\underline{u} + \underline{\mu}$$

where $\underline{\mu}$ is a column vector of constants. $\underline{y}$ is a column vector formed from $\{y_j\}_1^m$; $\underline{u}$ is the column vector formed from $\{u_j\}_1^m$ and G is a non-singular (mxm) matrix of constants has the property that

$$\underline{\alpha}'\underline{y} = (\alpha'G)\underline{u} + \underline{\alpha}'\underline{\mu}$$

is a linear combination of independent normal variates for any $\underline{\alpha}$ (since $\underline{\alpha}'G \neq \underline{0}$). Clearly then the probability density on $\underline{y}$ will satisfy our requirements whenever $|G| \neq 0$. We note that without loss of generality, we may assume $\{\mathcal{E}(u_j) = 0\}_{j=1}^{m}$; $\{\mathcal{E}(u_j^2) = 1\}_{j=1}^{m}$. Now since

$$f(u_1,u_2,\ldots,u_n) = \frac{1}{(2\pi)^{\frac{1}{2}m}} \exp - \tfrac{1}{2}\sum u_i{}^2$$

$$= \frac{1}{(2\pi)^{\frac{1}{2}m}} \exp - \tfrac{1}{2}\underline{u}'\underline{u}$$

we can find the density on $\{y_j\}_1^m$. the Jacobian of the linear transformation from $\{u_j\}_1^m$ to $\{y_j\}_1^m$ is simply $\|G\|^{-1}$ and hence

$$f(y_1,y_2,\ldots,y_m) = \frac{\|G\|^{-1}}{(2\pi)^{\frac{1}{2}m}} \exp - \tfrac{1}{2}(\underline{y}-\underline{\mu})'(GG')^{-1}(\underline{y}-\underline{\mu}).$$

Since V (say) defined by V=GG' is positive definite and since the set of all non-singular G will develop the set of all positive definite V we see that

$$f(y_1,y_2,\ldots y_m) = \frac{|V|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}m}} \exp - \tfrac{1}{2}(\underline{y}-\underline{\mu})'V^{-1}(\underline{y}-\underline{\mu})$$

should have the property that any linear combination of the $\{y_j\}$(say $\underline{\alpha}'\underline{y}$, for a set of constants $\underline{\alpha}(\neq\underline{0})$) has a non-singular normal univariate density.

In Chapter 2 of these notes we will forget the origins of the density on $\{y_j\}_{j=1}^m$ saying "here is a function of the m variables $y_1,y_2,\ldots,y_m$"; and ask "is it a valid probability density" and "what properties does the surface have?"

Observe that, by virtue of its derivation, the surface given above does have the original required properties. It is natural to inquire whether this be the only surface having the property that $\underline{\alpha}'\underline{y}$ has a non-singular univariate normal density for any $\underline{\alpha}(\underline{\alpha}\neq\underline{0})$. It develops that this is the case and this problem will be dealt with later.

1.2 Notation

It is almost certainly true that the most fearsome feature of a first encounter with multivariate theory is the notation. The notation and development is greatly facilitated by a liberal use of the matrix algebra; even so, without a few consistencies in notation the reader may easily loose the thread of a proof. It would be of benefit then to memorize these few consistencies:

An underscored lower case letter is a column vector; thus, $\underline{x},\underline{y},\underline{a},\underline{\alpha},\underline{\beta}$ are column vectors.

In almost all cases greek letters are used as constants, the exception to this is the symbol χ^2 which will have its universal meaning. The other exception to this rule is the use of "eta" (η) as a random variable.

The symbols x,y,z,u,w,t and η have been used to represent random variables; all other roman letters are inevitably constants.

Upper case letters (greek and roman) represent matrices. These upper case letters are not underscored.

V is inevitably a positive definitive matrix.

It is frequently desirable to form a matrix of a rectangular array and write $A = (a_{ij})$. It is to be understood that the element a_{ij} appears in the i-th row and j-th column of A. If C be a matrix it is sometimes required to draw attention to a particular element, row or column of C so that

$(C)_{ij}$ will be understood to mean the element in the i-th row, j-th column of C; sometimes when a series of such elements are required, it is more convenient to resort to the use of the corresponding lower case letter and use c_{ij} for $(C)_{ij}$.

$(C)_{i.}$ will be understood to mean the i-th row of C is used to form the row vector $(C)_{i.}$.

$(C)_{.j}$ will be understood to mean the j-th column of C is used to form the column vector $(C)_{.j}$.

The symbols P and Q are held to be orthogonal matrices throughout. Some other symbols may be orthogonal in a particular section.

The symbol m is used only in the context of a dimension of a vector of random variables or a matrix of random variables. Symbols p and q will represent throughout partitions of m, that is, m = p+q.

It is frequently required to partition vectors and matrices in the development of the theory. We will write for example:

$$\underline{X} = \left[\begin{array}{c} \underline{X}_{(1)} \\ \hline \underline{X}_{(2)} \end{array}\right]$$

and

$$V = \left[\begin{array}{c|c} V_{11} & V_{12} \\ \hline V_{21} & V_{22} \end{array}\right]$$

The meaning of

$\underline{x}_{(1)}$: an upper partition of $\underline{x}$;

x_1 : the topmost element in $\underline{x}$ and

$\underline{x}_1$: the first vector in a set $\{\underline{x}_j\}$ of vectors

should be contrasted.

The meaning of

V_{11} : the top-left partition in V should be contrasted with
$(V)_{11}$ the top-left element in V.

CHAPTER 2

THE MULTIVARIATE NORMAL DENSITY

2.1 Introduction

In this chapter we define the multivariate normal density and investigate some of its properties. The more elementary properties and definitions concerning multivariate probability density functions are assumed known; for example, the definitions of marginal density, conditional densities and independence. It is thought that a familiarity with the results of Appendix A will be of help to the reader.

2.2 The multivariate normal density.

The multivariate normal density function on variables $x_1, x_2, \ldots, x_m$ is defined over the entire m-dimensional Euclidean space and has the form

$$(2.2.1) \qquad f(x_1, x_2, \ldots, x_m) = \frac{|\Lambda|^{+\frac{1}{2}}}{(2\pi)^{\frac{1}{2}m}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'\Lambda(\underline{x}-\underline{\mu})}$$

with

$$(2.2.2) \qquad \underline{x}' = (x_1, x_2, \ldots, x_m)$$

a vector of random variagles; or a point in m-dimensional Euclidean space

$$(2.2.3) \qquad \underline{\mu}' = (\mu_1, \mu_2, \ldots, \mu_m)$$

a vector of constants and

$$(2.2.4) \qquad \Lambda = \Lambda' = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix}$$

a symmetric matrix of constants.

The existence of $|A|$ in (2.2.1) implies that it be non-singular. If $f(x_1, x_2, \ldots, x_m)$ is to be a density function, then the integral of the density function over the entire m-dimensional Euclidean space must be unity; if it is to be finite then, of necessity

$$(2.2.5) \qquad (\underline{x}-\underline{\mu})'A(\underline{x}-\underline{\mu}) \geq 0, \quad \text{for all } \underline{x}.$$

Since A is non-singular and the associated quadratic form in (2.2.5) is non-negative, it follows that A must be positive definite. Conversely, if A is positive definite, then A will be non-singular and (2.2.5) will hold (with equality if and only if $\underline{x} = \underline{\mu}$).

There exists a nonsingular matrix M such that

$$(2.2.6) \qquad M'AM = I.$$

Upon making the linear transformation $\underline{x}-\underline{\mu} = M\underline{y}$, the Jacobian of the transformation being $|M| = |A|^{-\frac{1}{2}}$, we see that

$$(2.2.7) \qquad \underset{-\infty<x_j<\infty}{\int\cdots\int} \frac{|A|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}m}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'A(\underline{x}-\underline{\mu})} dx_1\ldots dx_m$$

$$= \underset{-\infty<y_i<\infty}{\int\cdots\int} \frac{e^{-\frac{1}{2}\sum_{i=1}^{m} y_i^2}}{(2\pi)^{\frac{1}{2}m}} dy_1\ldots dy_m = 1$$

Thus for positive definite A, the density expressed in (2.2.1) is a true probability density function.

The properties of $f(x_1, x_2, \ldots, x_m)$, for example the moments and marginal densities, are more conveniently expressed in terms of the inverse of A so that we will rewrite (2.2.1) in a slightly different form:

$$(2.2.8)\qquad f(x_1,x_2,\ldots,x_m) = \frac{|V|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}m}}\, e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'V^{-1}(\underline{x}-\underline{\mu})}$$

A convenient short-hand notation will be adopted. If m variables $x_1, x_2, \ldots, x_m$ have the density expressed in (2.2.8) then we shall write simply

$$(2.2.9)\qquad \underline{x} \frown N_m(\underline{\mu}, V)$$

which is to be read "$\underline{x}$ has an m-variate normal density with mean vector $\underline{\mu}$ and dispersion matrix V."

At the moment, the terms "mean vector' and "dispersion matrix" are words to describe sets of constants (or parameters) occurring in the density function. We shall show later that

$$(2.2.10)\qquad \mathcal{E}(\underline{x}) = \begin{bmatrix} \mathcal{E}(x_i) \\ \mathcal{E}(x_2) \\ \vdots \\ \mathcal{E}(x_m) \end{bmatrix} = \underline{\mu}$$

and that

$$(2.2.11)\qquad \mathcal{E}(\underline{x}-\underline{\mu})(\underline{x}-\underline{\mu})' = \begin{bmatrix} \mathrm{Var}(x_1) & \mathrm{cov}(x_1,x_2) & \ldots & \mathrm{cov}(x_1,x_m) \\ \mathrm{cov}(x_2,x_1) & \mathrm{Var}(x_2) & \ldots & \mathrm{cov}(x_2,x_m) \\ \vdots & \vdots & & \vdots \\ \mathrm{cov}(x_m,x_1) & \mathrm{cov}(x_m,x_2) & \ldots & \mathrm{Var}(x_m) \end{bmatrix} = V$$

where $\mathrm{Var}(x_i)$ is the variance of x_i and $\mathrm{cov}(x_i,x_j)$ is the covariance between x_i and x_m. (For this reason the dispersion matrix is sometimes referred to as the variance-covariance matrix.)

Theorem 1. (for a slight extension, see excercise 5).

If $\underline{x} \frown N_m(\underline{\mu}, V)$ and C is an (mxm), non-singular matrix, then $\underline{y} = C\underline{x} \frown N_m(C\underline{\mu}, CVC')$.

The Jacobian of the transformation $\underline{x} = C^{-1}\underline{y}$ is the modulus of $|C|^{-1}$. Hence

$$(2.2.12) \qquad f(y_1,y_2,\dots,y_m) = \frac{\mathrm{mod}|C|^{-1}|V|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}m}} e^{-\frac{1}{2}(C^{-1}\underline{y}-\underline{\mu})'V^{-1}(C^{-1}\underline{y}-\underline{\mu})}$$

or, by simple rearrangement

$$(2.2.13) \qquad f(y_1,y_2,\dots,y_m) = \frac{|CVC'|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}m}} e^{-\frac{1}{2}(\underline{y}-C\underline{\mu})'(CVC')^{-1}(\underline{y}-C\underline{\mu})}$$

so that, in our short-hand notation, $\underline{y} \frown N_m(C\underline{\mu}, CVC')$.

2.3 The marginal densities.

The marginal density of the set of variables $x_1, x_2, \dots, x_p$ is defined as

$$(2.3.1) \qquad f(x_1,x_2,\dots,x_p) = \int \dots \int f(x_{.},x_2,\dots,x_m)dx_{p+1},dx_{p+2},\dots,dx_m$$

$$\{-\infty<x_i<\infty\}_{p+1}^{m}$$

Suppose the random vector $\underline{x}'$ be partitioned into two vectors $\underline{x}'_{(1)} = (x_1,x_2,\dots,x_p)$ and $\underline{x}'_{(2)} = (x_{p+1},x_{p+2},\dots,x_m)$; that is

$$(2.3.2) \qquad \underline{x} = \begin{bmatrix} \underline{x}_{(1)} \\ \cdots \\ \underline{x}_{(2)} \end{bmatrix}.$$

The mean vector and dispersion matrix are correspondingly partitioned

$$(2.3.3) \qquad \underline{\mu} = \begin{bmatrix} \underline{\mu}_{(1)} \\ \cdots \\ \underline{\mu}_{(2)} \end{bmatrix} .$$

$$(2.3.4) \qquad V = \left[\begin{array}{c:c} V_{11} & V_{12} \\ \hdashline V_{21} & V_{22} \end{array}\right]$$

with V_{11} a (pxp) symmetric matrix and $V_{12} = V'_{21}$ a $(p \times \overline{m-p})$ matrix.

Theorem 2.

If $\underline{x} \frown N_m(\underline{\mu}, V)$ then with the partitions as defined in equations (2.3.2), (2.3.3), and (2.3.4), $\underline{x}_{(1)} \frown N_p(\underline{\mu}_{(1)}, V_{11})$

Let $A = V^{-1}$, the inverse of the partitioned matrix V, and partition A in the same manner as V, then

$$(2.3.5) \qquad f(x_1, x_2, \ldots, x_m) = \frac{|A|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}m}} e^{-\frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}(\underline{x}_{(1)}-\underline{\mu}_{(i)})'A_{ij}(\underline{x}_{(j)}-\underline{\mu}_{(j)})}$$

But

$$(2.3.6) \qquad \sum_{i=1}^{2}\sum_{j=1}^{2}(\underline{x}_{(i)}-\underline{\mu}_{(i)})'A_{ij}(\underline{x}_{(j)}-\underline{\mu}_{(j)}) =$$

$$(\underline{x}_{(1)}-\underline{\mu}_{(1)})'(A_{11}-A_{12}A_{22}^{-1}A_{21})(\underline{x}_{(1)}-\underline{\mu}_{(1)}) \quad +$$

$$[(\underline{x}_{(2)}-\underline{\mu}_{(2)})+A_{22}^{-1}A_{21}(\underline{x}_{(1)}-\underline{\mu}_{(1)})]'A_{22}[(\underline{x}_{(2)}-\underline{\mu}_{(2)}+A_{22}^{-1}A_{21}(\underline{x}_{(1)}-\underline{\mu}_{(1)})]$$

Making the transformation from $\underline{x}_{(2)}$ to $\underline{y}$ defined by

(2.3.7) $$\underline{y} = \underline{x}_{(2)}-\underline{\mu}_{(2)} + A_{22}^{-1}A_{21}(\underline{x}_{(1)}-\underline{\mu}_{(1)})$$

and using the results that

(2.3.8) $$|A| = |A_{22}||A_{11}-A_{12}A_{22}^{-1}A_{21}| = |A_{22}||V_{11}|^{-1}$$

(2.3.9) $$V_{11}^{-1} = (A_{11}-A_{12}A_{22}^{-1}A_{21})$$

(See Appendix A),

we have

(2.3.10) $$f(x_1,x_2,\ldots,x_p;\ y_1,y_2,\ldots,y_{m-p})$$

$$= \frac{|V_{11}|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}p}} e^{-\frac{1}{2}(\underline{x}_{(1)}-\underline{\mu}_{(1)})'V_{11}^{-1}(\underline{x}_{(1)}-\underline{\mu}_{(1)})} \quad X$$

$$\frac{|A_{22}|^{+\frac{1}{2}}}{(2\pi)^{\frac{1}{2}(m-p)}} e^{-\frac{1}{2}\underline{y}'A_{22}\underline{y}}$$

and

$$\begin{cases} -\infty < x_i < \infty & i = 1, 2, \ldots, p \\ -\infty < y_i < \infty & i = 1, 2, \ldots, m-p. \end{cases}$$

The second factor in the product is the density $N_{m-p}(\underline{0}, A_{22}^{-1})$ and as such will integrate to unity over $y_1, y_2, \ldots, y_{m-p}$ so that

$$(2.3.11) \qquad f(x_1,x_2,\ldots,x_p) = \frac{|V_{11}|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}p}} e^{-\frac{1}{2}(\underline{x}_{(1)}-\underline{\mu}_{(1)})'V_{11}^{-1}(\underline{x}_{(1)}-\underline{\mu}_{(1)})}$$

and the theorem is proved.

Using the fact that $V_{21}V_{11}^{-1} = -A_{22}^{-1}A_{21}$ we observe the following corollary; that with $\underline{x}_{(1)}$ and $\underline{x}_{(2)}$ defined as in the theorem, then $\underline{x}_{(2)}-\underline{\mu}_{(2)}-V_{21}V_{11}^{-1}(\underline{x}_{(1)}-\underline{\mu}_{(1)})$ has the density $N_{m-p}(\underline{0}, V_{22}-V_{21}V_{11}^{-1}V_{12})$ and is independent of $\underline{x}_{(1)}$.

Theorem 3.

If $\underline{x} \frown N_m(\underline{\mu}, V)$ then $\mathcal{E}(x_\alpha) = \mu_\alpha$; $\mathrm{Var}(x_\alpha) = V_{\alpha\alpha}$ and $\mathrm{Cov}(x_\alpha, x_\beta) = V_{\alpha\beta}$.

Putting p = 2 in Theorem 2, equation (2.3.11) may be written in the form

$$(2.3.12) \qquad f(x_1,x_2) = \frac{e^{-\frac{1}{2}\left[\frac{(x_1-\mu_1)^2}{V_{11}} - \frac{2V_{12}(x_1-\mu_1)(x_2-\mu_2)}{V_{11}V_{22}} + \frac{(x_2-\mu_2)^2}{V_{22}}\right]\left(1-\frac{V_{12}^2}{V_{11}V_{22}}\right)^{-1}}}{2\pi\sqrt{V_{11}V_{22}-V_{12}^2}}$$

this may be identified with the normal bivariate density and the results of the theroem follow.

The results of Theorems 2 and 3 also obtain from the joint characteristic function of $x_1, x_2, \ldots, x_p$ which is to be considered in a later chapter.

2.4 The conditional densities.

The density of the set $x_{p+1}, x_{p+2}, \ldots, x_m$ conditional on $x_1, x_2, \ldots, x_p$ which will be written

$$f(x_{p+1}, x_{p+2}, \ldots, x_m | x_1, x_2, \ldots, x_p)$$

is defined by

$$f(x_{p+1}, x_{p+2}, \ldots, x_m | x_1, x_2 \ldots, x_p) = \frac{f(x_1, x_2, \ldots, x_m)}{f(x_1, x_2, \ldots, x_p)}. \tag{2.4.1}$$

Theorem 4.

If $\underline{x} \frown N_m(\underline{\mu}, V)$, then with the partitions as defined in equations 2.3.2), (2.3.3), and (2.3.4),

$$*\underline{x}_{(2)} | \underline{x}_{(1)} \frown N_{m-p}(\underline{\mu}_{(2)} + V_{21}V_{11}^{-1}(\underline{x}_{(1)} - \underline{\mu}_{(1)});\ V_{22} - V_{21}V_{11}^{-1}V_{12})$$

The result follows immediately using the corollary to Theorem 2. Since $\underline{x}_{(2)} - \underline{\mu}_{(2)} - V_{21}V_{11}^{-1}(\underline{x}_{(1)} - \underline{\mu}_{(1)})$ is independent of $\underline{x}_{(1)}$, the density of this $(\overline{\overline{m-p}}\text{x}1)$ vector for fixed $\underline{x}_{(1)}$ is simply $N_{m-p}(\underline{0}, V_{22} - V_{21}V_{11}^{-1}V_{12})$. The theorem is proved after making the

---

*to be read "$\underline{x}_{(2)}$ conditional on $\underline{x}_{(1)}$ has the density . . ."

the change in location.

The mean vector for $\underline{x}_{(2)}$ conditional on $\underline{x}_{(1)}$ is a linear function of the elements of $\underline{x}_{(1)}$; we may write

$$(2.4.2) \quad \mathcal{E}(x_i \mid x_1, x_2, \ldots, x_p) = \mu_i + \sum_{j=1}^{p} \beta_{ij.1,2,\ldots,j-1,j+1,\ldots,p}(x_j - \mu_j) \qquad i = p+1, p+2, \ldots, m$$

so that $\beta_{ij.1,2,\ldots,j-1,j+1,\ldots,p}$ expresses the expected change in x_i per unit change in x_j when the set $\{x_\ell\}_{\ell=1;\ \ell \neq j}^{p}$ is held fixed.

The conditional dispersion matrix (matrix of partial variances and covariances) $V_{22.(1)}$ defined by

$$(2.4.3) \quad V_{22.(1)} = V_{22} - V_{21}V_{11}^{-1}V_{12}$$

has for its (st)-th element, the covariance between x_t and x_s when the vector $(x_1, x_2, \ldots, x_p)$ is held fixed. Giving this covariance the symbol $v_{st.(1)}$, we see that

$$(2.4.4) \quad v_{st.(1)} = v_{st} - (V_{21})_{s.} V_{11}^{-1} (V_{12})_{.t} \qquad {s \atop t} = p+1, p+2, \ldots, m.$$

Since V_{11}^{-1} is positive definite, it follows at once that

$$(2.4.5) \quad v_{ss.(1)} \leq v_{ss}$$

with the equality if and only if $(V_{21})_{s.} = (V_{12})'_{.s} = \underline{0}'$.

The partial correlations are defined by

$$(2.4.6) \quad \rho_{st.(1)} = v_{st.(1)} \Big/ \sqrt{v_{ss.(1)} v_{tt.(1)}}\ .$$

For other results connecting correlations with partial correlations the reader is referred to Exercises to Chapter 2.

2.5 The canonical correlation coefficient.

The vector of random variables, $\underline{x} \frown N_m(\underline{\mu}, v)$ is partitioned into $\underline{x}' = [\underline{x}'_{(1)} \vdots \underline{x}'_{(2)}]$, the vector $\underline{x}_{(1)}$ having p elements. The mean vector $\underline{\mu}$ and dispersion matrix V are partitioned to conform with the partitioning of $\underline{x}$.

The scalar quantities $\underline{\alpha}_1'\underline{x}_{(1)}$ and $\underline{\alpha}_2'\underline{x}_{(2)}$ have a normal bivariate density, in fact

$$(2.5.1) \qquad \begin{bmatrix} \underline{\alpha}_1'\underline{x}_{(1)} \\ \underline{\alpha}_2'\underline{x}_{(2)} \end{bmatrix} \frown N_2\left(\begin{bmatrix} \underline{\alpha}_1'\mu_{(1)} \\ \underline{\alpha}_2'\underline{\mu}_{(2)} \end{bmatrix} ; \begin{bmatrix} \underline{\alpha}_1'V_{11}\underline{\alpha}_1 & \underline{\alpha}_1'V_{12}\underline{\alpha}_2 \\ \underline{\alpha}_2'V_{21}\underline{\alpha}_1 & \underline{\alpha}_2'V_{22}\underline{\alpha}_2 \end{bmatrix} \right)$$

The correlation between $\underline{\alpha}_1'\underline{x}_{(1)}$ and $\underline{\alpha}'_{(2)}\underline{x}_{(2)}$ is given by

$$(2.5.2) \qquad \mathrm{Corr}(\underline{\alpha}_1'\underline{x}_{(1)};\underline{\alpha}_2'\underline{x}_{(2)}) = \underline{\alpha}_1'V_{12}\underline{\alpha}_2 \Big/ (\underline{\alpha}_1'V_{11}\underline{\alpha}_1)^{1/2}(\underline{\alpha}_2'V_{22}\underline{\alpha}_2)^{1/2}$$

The canonical correlation, $\rho_{(1).(2)}$, between the vectors $\underline{x}_{(1)}$ and $\underline{x}_{(2)}$ is defined by

$$(2.5.3) \qquad \rho_{(1).(2)} = \max_{\{\underline{\alpha}_1\underline{\alpha}_2\}} \mathrm{Corr}(\underline{\alpha}_1'\underline{x}_{(1)};\underline{\alpha}_2'\underline{x}_{(2)})$$

Since the right hand side of (2.5.2) is independent of the modulus of the $\underline{\alpha}_i$ we may seek to maximize $\underline{\alpha}_1'V_{12}\underline{\alpha}_2$ subject to $\underline{\alpha}_1'V_{11}\underline{\alpha}_1 = \underline{\alpha}_2'V_{22}\underline{\alpha}_2 = 1$

with Q defined by

$$(2.5.4) \qquad Q = 2\underline{\alpha}_1'V_{12}\underline{\alpha}_2 - \lambda_1(\underline{\alpha}_1'V_{11}\underline{\alpha}_1-1) - \lambda_2(\underline{\alpha}_2'V_{22}\underline{\alpha}_2-1),$$

the stationary values of corr$(\underline{\alpha}_1'\underline{x}_{(1)};\underline{\alpha}_2'\underline{x}_{(2)})$ are given by the solutions of

$$\frac{dQ}{d\underline{\alpha}_1} = \underline{0}; \quad \frac{dQ}{d\underline{\alpha}_2} = \underline{0}; \quad \frac{dQ}{d\lambda_1} = \frac{dQ}{d\lambda_2} = 0. \tag{2.5.5}$$

that is, by the solutions of

$$\left\{\begin{aligned} V_{12}\underline{\alpha}_2 - \lambda_1 V_{11}\underline{\alpha}_1 &= \underline{0} \\ V_{21}\underline{\alpha}_1 - \lambda_2 V_{22}\underline{\alpha}_2 &= \underline{0} \\ \underline{\alpha}_1' V_{11}\underline{\alpha}_1 &= 1 \\ \underline{\alpha}_2' V_{22}\underline{\alpha}_2 &= 1 \end{aligned}\right. \tag{2.5.6}$$

Premultiplying the first equation by $\underline{\alpha}_1'$ and the second by $\underline{\alpha}_2'$, it is found that $\lambda_1 = \lambda_2 = \rho_{(1).(2)}$. Eliminate $\underline{\alpha}_2$ from the first two equations to obtain

$$(V_{11}^{-1}V_{12}V_{22}^{-1}V_{21} - \rho^2_{(1).(2)}I)\underline{\alpha}_1 = 0. \tag{2.5.7}$$

Thus $\rho^2_{(1).(2)}$ is the maximum latent root of $V_{11}^{-1}V_{12}V_{22}^{-1}V_{21}$ with $\underline{\alpha}_1$ the associated eigen vector. Alternatively, $\rho^2_{(1).(2)}$ is the maximum latent root of $V_{22}^{-1}V_{21}V_{11}^{-1}V_{12}$ and $\underline{\alpha}_2$ the associated eigen vector. It is important to note that $\rho^2_{(1).(2)}$ is zero if and only if $V_{12} = (0)$, that is if $f(\underline{x}_{(1)},\underline{x}_{(2)}) = f(\underline{x}_{(1)})f(\underline{x}_{(2)})$.

The vectors $\underline{\alpha}_1$ and $\underline{\alpha}_2$ are determined by (2.5.7) and (2.5.8) apart from the sign unless the maximum root of $V_{11}^{-1}V_{12}V_{22}^{-1}V_{21}$ is repeated; in this case any $\underline{\alpha}_1,\underline{\alpha}_2$ satisfying (2.5.7) and (2.5.8) may be used. The sign of $\rho_{(1).(2)}$ can be determined by considering the leading element in the equation $V_{12}\underline{\alpha}_2 = \rho_{(1).(2)}V_{11}\underline{\alpha}_1$ (see (2.5.6)).

2.6 The multiple correlation coefficient.

The coefficient of multiple correlation is a special case of a canonical correlation coefficient. We again suppose $\underline{x} \frown N_m(\mu, V)$ and partition $\underline{x}$, $\underline{\mu}$ and V as in section 2.5. Suppose x_i is an element of $\underline{x}_{(1)}$; the scalar quantities x_i and $\underline{\alpha}'\underline{x}_{(2)}$ have a normal bivariate density, in fact

$$(2.6.1) \qquad \begin{bmatrix} x_i \\ \underline{\alpha}'\underline{x}_{(2)} \end{bmatrix} \frown N_2\left(\begin{bmatrix} \mu_i \\ \underline{\alpha}'\underline{\mu}_{(2)} \end{bmatrix} ; \begin{bmatrix} v_{ii} & V_{i2}\underline{\alpha} \\ \underline{\alpha}'V_{2i} & \underline{\alpha}'V_{22}\underline{\alpha} \end{bmatrix} \right)$$

v_{ii} is a scalar quantity; V_{22} is the $\overline{m-p}\times\overline{m-p}$ dispersion matrix associated with the random vector $\underline{x}_{(2)}$. $V_{i2} = V_{2i}$ is a row vector of the covariances between x_i and the row vector $(x_{p+1}, x_{p+2}, \ldots, x_m) = \underline{x}'_{(2)}$.

The correlation between x_i and $\underline{\alpha}'\underline{x}_{(2)}$ is given by

$$(2.6.2) \qquad \operatorname{Corr}(x_i, \underline{\alpha}'\underline{x}_{(2)}) = V_{i2}\underline{\alpha} \Big/ \sqrt{v_{ii}(\underline{\alpha}'V_{22}\underline{\alpha})} .$$

The multiple correlation $R_{i.(2)}$ (alternatively written $R_{i.p+1,p+2,\ldots,m}$ if it be required to stress the elements in $\underline{x}_{(2)}$) is defined by

$$(2.6.3) \qquad R_{i.(2)} = \max_{\underline{\alpha}} \operatorname{Corr}(x_i, \underline{\alpha}'\underline{x}_{(2)})$$

The proceedure is much the same as that for canonical correlation, the stationary values for corr$(x_i, \underline{\alpha}'\underline{x}_{(2)})$ being given by the solutions of

(2.6.4) $$\frac{dQ}{d\underline{\alpha}} = \frac{dQ}{d\lambda} = 0$$

where

(2.6.5) $$Q = 2V_{i2}\underline{\alpha} - \lambda(\underline{\alpha}'V_{22}\underline{\alpha}-1)$$

that is, by the solutions of

(2.6.6) $$\begin{cases} V_{2i} - \lambda V_{22}\underline{\alpha} = 0 \\ \underline{\alpha}'V_{22}\underline{\alpha} = 1 \end{cases}$$

Premultiplying the first equation by $\underline{\alpha}'$ and, bearing in mind that $\underline{\alpha}'V_{2i} = V_{2i}'\underline{\alpha}$, it is found that $\lambda = R_{i.(2)}$. Again, using the first equation, $\underline{\alpha}$ is a multiple of the vector $V_{22}^{-1}V_{2i}$ so that

(2.6.7) $$R_{i.(2)} = \left(\frac{V_{i2}V_{22}^{-1}V_{2i}}{v_{ii}}\right)^{\frac{1}{2}}$$

The sign of the square-root in equation (2.6.7) is determined by considering the leading element of the equation $V_{2i} = R_{i.(2)}V_{22}\underline{\alpha}$ (see 2.6.6).

This result is also obtainable from equations (2.5.7) and (2.5.8) after replacing suffix 1 by i; $\underline{\alpha}_1$ by unity; $\underline{\alpha}_2$ by $\underline{\alpha}$, and $\rho^2_{(1).(2)}$ by $R^2_{i.(2)}$.

It is important to note that $R_{i.(2)}$ is zero if and only if $V_{i2} = \underline{0}'$, that is if $\mathrm{Cov}(x_i,x_j) = 0$, $i<p$, $j=p+1,p+2,\ldots,m$ so that $f(x_i,\underline{x}_{(2)}) = f(x_i)f(\underline{x}_{(2)})$

2.7 The multivariate normal distribution.

It is the writer's wish to draw a line of difference between "distribution" and "density" (at least for the purpose of these notes) to allow for the fact that a vector of random variables may have a joint distribution but not a joint density. This is perhaps best illustrated by an example; suppose that $\underline{w} \frown N_m(0, I)$, that is $\{w_i\}_1^m$, is a set of independent standard normal variates, and let $y_i = w_i - \bar{w}$, $i = 1, 2, \ldots, m$, with $\bar{w}$ the arithmetic mean of the $\{w_i\}_1^m$, then $\underline{y} = (y_1, y_2, \ldots, y_m)'$ has what we shall think of a joint distribution but no joint density over the entire m-dimensional Euclidean space (see Section 2.2). In the case where a vector (or matrix) of variables has a distribution but not necessarily a density, we will write $\sim$ rather than $\frown$ and define the multivariate normal distribution as follows:

If $\underline{w} \frown N_m(\underline{0}, I)$ and $\underline{y}$ is defined by $\underline{y} = \underline{\mu} + G\underline{w}$ where $\underline{\mu}$ is a kxl vector of constants and G is a kxm matrix of constants, then we will write $\underline{y} \sim N_k(\underline{\mu}, GG')$ which is to be read "$\underline{y}$ is a set of linear combinations of normal independent variables." We note that

$$(2.7.1) \qquad \begin{cases} \mathcal{E}(y_i) = \mu_i \\ \mathrm{Var}(y_i) = (GG')_{ii} \\ \mathrm{Cov}(y_i, y_j) = (GG')_{ij} = (GG')_{ji} \end{cases}$$

The following theorems follow at once from the definition

Theorem 5.

If $\underline{y} \sim N_k(\underline{\mu};GG')$ and $\underline{z} = C\underline{y}$ with C an ℓxk matrix of constants, then $\underline{z} \sim N_\ell(C\underline{\mu};CGG'C')$.

Theorem 6.

If $y' = (\underline{y}'_{(1)} \vdots \underline{y}'_{(2)})$ where $\underline{y}$ is defined in Theorem 7 and $\underline{y}_{(1)}$ is an hx1 partition of $\underline{y}$, then $\underline{y}_{(1)} \sim N_h(\underline{\mu}_{(1)};(GG')_{11})$ with $\underline{\mu}_{(1)}$ and $(GG')_{11}$ the usual partitions of $\underline{\mu}$ and GG'.

The following theorem is a consequence of Theorem 1 in conjunction with the extension to that theorem given in exercise 5.,

Theorem 7.

If $\underline{y} \sim N_k(\underline{\mu};GG')$ and if $|GG'| \neq 0$, then $\underline{y} \frown N_k(\underline{\mu};GG')$.

We may define a conditional distribution as the distribution of one set of variables when another set is held fixed and give the following theorem.

Theorem 8.

Let $\underline{y} \sim N_k(\underline{\mu}; V)$ with V positive semi-definite (since it is of the form GG') and $\underline{y}' = (\underline{y}'_{(1)} \vdots \underline{y}'_{(2)})$ where $\underline{y}_{(1)}$ is an h×1 partition of $\underline{y}$. Suppose $\underline{\mu}$ and V are suitably partitioned and that V_{11}(h×h) is nonsingular and, therefore, positive definite, then

$$\underline{y}_{(2)}|\underline{y}_{(1)} \sim N_{k-h}(\underline{\mu}_{(2)}+V_{21}V_{11}^{-1}(y_{(1)}-\mu_{(1)}),\ V_{22}-V_{21}V_{11}^{-1}V_{12}).$$

Since V is positive semi-definite, there exists a lower triangular matrix T with real elements such that $V = TT'$. With $\underline{w} \sim N_m(0;I)$ we may write $\underline{y} = \underline{\mu}+T\underline{w}$, or using the partitions

$$(2.7.2) \qquad \begin{cases} \underline{y}_{(1)} = \underline{\mu}_{(1)} + T_{11}\underline{w}_{(1)} & (|T_{11}| \neq 0) \\ \underline{y}_{(2)} = \underline{\mu}_{(2)} + T_{21}\underline{w}_{(1)} + T_{22}\underline{w}_{(2)} & \end{cases}$$

Hence

$$(2.7.3) \qquad \underline{y}_{(2)} = \underline{\mu}_{(2)} + T_{21}T_{11}^{-1}(y_{(1)}-\mu_{(1)}) + T_{22}\underline{w}_{(2)}$$

so that, by definition and since $\underline{y}_{(1)}$ and $\underline{w}_{(2)}$ are independent

$$(2.7.4) \qquad \underline{y}_{(2)}|y_{(1)} \sim N_{k-h}\left(\mu_{(2)}+T_{21}T_{11}^{-1}(y_{(1)}-\mu_{(1)});\ T_{22}T_{22}'\right)$$

Now

$$(2.7.5) \qquad \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} = \left[\begin{array}{c|c} T_{11} & (0) \\ \hline T_{21} & T_{22} \end{array}\right] \left[\begin{array}{c|c} T_{11}' & T_{21}' \\ \hline (0) & T_{22}' \end{array}\right]$$

so that

$$(2.7.6)\qquad \begin{cases} V_{21}V_{11}^{-1} = T_{21}T_{11}^{-1} \\ V_{22}-V_{21}V_{11}^{-1}V_{12} = T_{22}T'_{22} \end{cases} .$$

Using (2.7.2) and (2.7.6) the proof is completed.
We may proceed to the definition of a cononical correlation as before and note the following theorems:

Theorem 9.

If $\underline{y} \sim N_k(\underline{\mu}, V)$ with V positive semi-definite such that $|V_{11}| \neq 0$ and $|V_{22}| \neq 0$, then $\rho^2_{(1).(2)} = 1$.

Using the work of Section 2.5, we have in this case that $\rho^2_{(1).(2)}$ is the maximum latent root of

$$V_{22}^{-1}V_{21}V_{11}^{-1}V_{12} \quad \text{but,}$$

$$(2.7.7)\qquad |V| = 0 = |V_{11}||V_{22}-V_{21}V_{11}^{-1}V_{12}|$$

that is

$$(2.7.8)\qquad |V_{11}||V_{22}||I-V_{22}^{-1}V_{21}V_{11}^{-1}V_{12}| = 0$$

since $|V_{11}| \neq 0$ and $|V_{22}| \neq 0$, it follows that $|V_{22}^{-1}V_{21}V_{11}^{-1}V_{12}-I| = 0$; that is $V_{22}^{-1}V_{21}V_{11}^{-1}V_{12}$ has a latent root of unity.

If $|V_{11}| = 0$ then for some $\underline{\lambda}$, $V_{11}\underline{\lambda} = \underline{0}$ so that the solution for $\underline{a}_1$ in equations (2.5.6) will not be unique; in fact, if $\underline{a}_1$ satisfies equation (2.5.6) so also does $\underline{a}_1 + c\underline{\lambda}$ for any c.

It is not the writer's wish to make a mountain out of the difference between a distribution and a density (as we have defined them in these notes). Nevertheless, it is often more convenient to consider theorems concerning variates which have a distribution but not necessarily a density; this comes out quite strongly when we consider Wishart variates (i.e., variates having a Wishart distribution) in the next chapter.

2.8 The distribution of a matrix of normal variables.

In the development of the multivariate theory, we occasionally need to consider the joint density of a matrix of normal variables.

Let X be an mxn matrix of variates, x_{ij}; that is

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \tag{2.8.1}$$

then if x_{ij} is normally distributed and

$$\mathcal{E}(x_{ij}) = \mu_{ij} \tag{2.8.2}$$

$$\operatorname{Cov}(x_{ij}, x_{pq}) = \Sigma_{(i-1)n+j;(p-1)n+q} \tag{2.8.3}$$

we will write $X \sim N_{m\times n}(\underline{\mu}, \Sigma)$ with $\underline{\mu}$ an mxn matrix of means and Σ an mnxmn matrix of covariances.

An immediate consequence of the definition of the covariance matrix is the following: if $(\underline{x}_j)_1^n$ is a set of independently distributed vectors such that $\underline{x}_j \sim N_n(\underline{\mu}_j, V)$ and if $X = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n)$ then $X \sim N(\underline{\mu}, V \otimes I)$ where $\underline{\mu} = (\underline{\mu}_1, \underline{\mu}_2, \ldots, \underline{\mu}_n)$ and $A \otimes B$ is the Kronecker (direct) proouct of A and B.

Theorem 10.

If $X \sim N_{m\times n}(\underline{\mu}, V \otimes U)$ then $Y = AXB' \sim N_{m\times n}\big((A\underline{\mu}B'; (AVA') \otimes (BUB')\big)$.

If $\Sigma = V \otimes U$ then $\Sigma_{(i-1)n+j:(p-1)n+q} = V_{ip}U_{jq}$, that is $\mathrm{Cov}(x_{ij}, x_{pq}) = V_{ip}U_{jq}$. Now by the definition of Y,

$$Y_{\alpha\beta} = \sum_i \sum_j a_{\alpha i}x_{ij}b_{\beta j} \text{ so that}$$

$$\mathcal{E}(Y_{\alpha\beta}) = \sum_i \sum_j a_{\alpha i}\mu_{ij}b_{\beta j} = (A\underline{\mu}B')_{\alpha\beta}, \text{ also}$$

$$\begin{aligned}\mathrm{Cov}(Y_{\alpha\beta}, Y_{\gamma\delta}) &= \sum_i \sum_j \sum_p \sum_q a_{\alpha i}b_{\beta j}a_{\gamma p}b_{\delta q}V_{ip}U_{jq}\\ &= (AVA')_{\alpha\gamma}(BUB')_{\beta\delta}\\ &= ((AVA') \times (BUB'))_{(\alpha-1)n+\beta:(\gamma-1)n+\delta}\end{aligned}$$

Finally since $y_{\alpha\beta}$ is a linear combination of normal variables, it is normally distributed so that we may write $AXB' = Y \sim N_{m\times n}(A\mu B', (AVA') \otimes (BUB'))$. An important consequence of this is the following result.

Theorem 11.

If $(\underline{x}_j)_1^n$ is a set of normally independently distributed vectors each with dispersion matrix V and mean vectors $(\underline{\mu}_j)_1^n$ and if

$$(\underline{y}_1,\underline{y}_2,\ldots,\underline{y}_n) = (\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n)P$$

then $(\underline{y}_j)_1^n$ are a set of normally independently distributed vectors each with dispersion matrix when P is orthogonal. If $\mathcal{E}(\underline{y}_j) = \underline{\xi}_j$, then $(\mu_j)_1^n$ and $(\underline{\xi}_j)_1^n$ are related by

$$(\underline{\xi}_1,\underline{\xi}_2,\ldots,\underline{\xi}_n) = (\underline{\mu}_1,\underline{\mu}_2,\ldots,\underline{\mu}_n)P.$$

The result is a direct consequence of Theorem 10 with U replaced by I, A replaced by I and B' replaced by P.

CHAPTER 3

THE WISHART DISTRIBUTION

3.1 Introduction

In this chapter, we define the Wishart variate and relate it to the distribution of certain generalized sums of squares. The density of the central Wishart matrix is obtained.

3.2 Definition of the Wishart Distribution.

It will be recalled that if $z_j \frown N_1(\mu_j, \sigma^2)$ and $(z_j)_1^\nu$ are independent, then $\sum z_j^2$ has a $\sigma^2\chi^2_\nu$ density with a noncentrality parameter proportional to $\sum_1^\nu \mu_j^2$. This may be thought of as a particular case of the Wishart random matrix wherein the matrix is 1x1, that is, scalar.

Let now $(\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_\nu) \frown N_{m\times\nu}((\underline{\mu}_1, \underline{\mu}_2, \ldots, \underline{\mu}_\nu); V \otimes I)$; that is, $(\underline{z}_j)_1^\nu$ are a set of independent, normally distributed random vectors each with dispersion matrix V and means given by $\mathcal{E}(\underline{z}_j) = \underline{\mu}_j$ then

$$W = \sum_{j=1}^{\nu} \underline{z}_j \underline{z}_j' \tag{3.2.1}$$

is a symmetric mxm matrix of random variables. We shall say that this random matrix has a Wishart distribution, dimension m, with degrees of freedom ν (a positive integer), dispersion matrix V and noncentrality matrix $\sum \underline{\mu}_j \underline{\mu}_j'$ adopting the short-hand notation $W \sim \mathcal{W}_m(V; \nu; \sum \underline{\mu}_j \underline{\mu}_j')$. If $\sum \underline{\mu}_j \underline{\mu}_j' = (0)$; we shall describe the distribution as "central". We note further that if $v \geq m$, then the matrix W is non-singular with probability one; in this event we shall show that the $\frac{1}{2}m(m+1)$ mathematically independent elements of W have a well-defined density. On the other hand, if $v \leq m-1$, then W is inevitably singular and the number of mathematically independent elements of W is $\frac{1}{2}v(v+1)$; we shall in this case describe the distribution of W as singular and use $\sim$ ("is distributed as") rather than $\frown$ ("has the density").

An immediate consequence of the definition of the Wishart distribution is

Theorem 12.

If $W_{(1)}$ and $W_{(2)}$ have independent distributions such that $W_{(i)} \sim \mathcal{W}_m(V;\ v_i : \Lambda_i)$, then $W_{(1)}+W_{(2)} \sim \mathcal{W}_m(V; v_1+v_2 : \Lambda_1+\Lambda_2)$.

Notice that each diagonal element, w_{ii}, say of W has a χ^2-distribution with v degrees of freedom and noncentrality parameter proportional to $(\sum \underline{\mu}_j \underline{\mu}_j')_{ii}$. The m χ^2-variates which are the diagonal elements of W are however, in general, correlated.

A consequence of Theorem 1 and the definition of the Wishart distribution is:

Theorem 13.

If $W \sim \mathcal{W}_m(V; v : \sum \underline{\mu}_j \underline{\mu}_j')$ then the distribution of CWC' is as $\mathcal{W}_h(CVC'; v : C\sum \underline{\mu}_j \underline{\mu}_j' C')$. Here we will require C to be rectangular h×m, possibly square, but certainly of full rank $h \leq m$.

We now turn to the extension to multivariate theory of the role of the idempotent matrix in the theory of least squares.

3.3 The generalized idempotent quadratic form.

It will be recalled that if $(x_j)_1^n$ are a set of homoscedastic independent normal variates with means given by the members of $(\mu_j)_1^n$ and if A is an nxn symmetric idempotent matrix of rank r, then

(3.3.1) $$\sum_i \sum_j a_{ij} x_i x_j \sim \sigma^2 \chi'^2_r\left(\frac{1}{2\sigma^2} \sum_i \sum_j a_{ij} \mu_i \mu_j\right)$$

with σ^2 the common variance of the $(x_j)_1^n$. We may write the quadratic form in (3.3.1) as $(x_1,x_2,x_3,\ldots,x_n)A(x_1,x_2,\ldots,x_n)'$ and think of the form where the scalar x_j is replaced by a vector $\underline{x}_j$ as a generalized quadratic form. If A is idempotent of rank r we will call the form $(\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n)A(\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n)'$ a generalized idempotent quadratic form of rank r and dimension m, the dimension being the number of elements in each $\underline{x}_j$.

Theorem 14.

If $X = (\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n) \frown N_{m\times n}(\Xi;\ V \otimes I)$ and A is idempotent of rank r, then $XAX' \sim \mathcal{W}_m(V;r:\Xi A \Xi')$.

Since A is idempotent, there exists an orthogonal nxn matrix P such that $PAP' = \Sigma$ with Σ a diagonal matrix $(\Sigma)_{ii} = 0$ or 1 and $\mathrm{tr}\Sigma = r$. Write $Y = XP'$, then $Y \frown N_{m\times n}(\Xi P';\ V \otimes I)$ and $XAX' = YPAP'Y' = Y\Sigma Y' = \sum_i (\Sigma)_{ii} \underline{y}_i \underline{y}_i'$. But, $\underline{y}_i \frown N_m(\underline{\xi}_i;V)$ and are independent where $\Xi = (\underline{\xi}_1 \underline{\xi}_2,\ldots,\underline{\xi}_n) = \Xi P'$. Finally by definition

$XAX' \sim \mathcal{W}_m(V;r:\xi A\xi')$; since $\sum_i (E)_{ii}\underline{\xi}_i\underline{\xi}_i' = \Xi E\Xi' = \xi P'EP\xi' = \xi A\xi'$, this completes the proof and the similarity with the case $m=1$ (least squares theory) should be noted. We should naturally expect many of the theorems connecting the idempotent quadratic form to the distribution theory under assumptions of normality to hold and in fact, they do with the Wishart distribution taking the place of the χ^2-distribution. The most important theorem is

Theorem 15.

Let $(A_t)_1^k$ be a set of nxn symmetric matrices such that any two of (a), (b), (c) hold

(a) A_t is idempotent, $t = 1, 2, \ldots, k$.

(b) $A_1+A_2\ldots+A_k$ is idempotent

(c) $A_tA_s = (0)$, $t \neq s$, $t,s = 1, 2, \ldots, k$.

then, if $X = (\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n) \frown N_{mxn}(\xi;\ V \times I)$ and $W_t = XA_tX'$, then $(W_t)_1^k$ are independently distributed such that $W_t \sim \mathcal{W}_m(V;r_t:\xi A_t\xi')$ where r_t is the rank of A_t.

If any two of (a), (b), (c) hold, then all three hold (theory of the idempotent matrix; see, for example, Graybill). By (a) $W_t \sim \mathcal{W}_m(V;r_t:\xi A_t\xi')$ directly from Theorem 14. It remains to show the independence of the $(W_t)_1^k$.

Using (c), there exists orthogonal P such that $PA_tP' = E_t$ $t = 1, 2, \ldots, k$ where E_t is a diagonal matrix such that $(E_t)_{ii}=0$ or 1 and $\mathrm{tr}E_t = r_t$. Since $A_tA_s = P'E_tE_sP = (0)$, $s\neq t$, we have that $E_t E_s = (0)$ $s \neq t$ so that no two matrices of the set $\{E_t\}_1^k$ have a non-zero diagonal element in common. With the transformation $Y = (\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n) = XP'$ we have $W_t \sim \sum_i (E_t)_{ii} \underline{y}_i \underline{y}_i'$ so that the W_t are linear combinations of different and independently distributed $\underline{y}_i \underline{y}_i'$; consequently, the $\{W_t\}_1^k$ are mutually independent. This completes a proof of the theorem.

A last theorem in this section concerns the independence of a linear function of normally distributed vectors and a generalized quadratic form in those vectors:

Theorem 16.

Let $X = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n) \frown N_{m \times n}(\mathcal{M}; V \otimes I)$ and suppose A to be an idempotent matrix of rank r. The linear combination, $\sum_{j=1}^{n} a_j \underline{x}_j$; of the $(\underline{x}_j)_1^n$ and the generalized idempotent quadratic form $\sum_i \sum_j a_{ij} \underline{x}_i \underline{x}_j'$ are distributed independently if and only if $A\underline{a} = \underline{0}$ where $A = (a_{ij})$ and $\underline{a}' = (a_1, a_2, \ldots, a_n)$.

The proof follows after the general pattern of this section and is left as an exercise. It should be noted that

$$\sum_1^n a_j \underline{x}_j = X\underline{a}$$

has a normal m-variate density, in fact

$$\sum_1^n a_j \underline{x}_j \frown N_m(\mathcal{M}\underline{a}, (\underline{a}'\underline{a})V) \tag{3.3.2}$$

where of course $\underline{a}'\underline{a}$ is a positive scalar quantity.

3.4 The Wishart Density (Integer degrees of freedom).

The symmetric matrix of random variables W has been defined by

$$(3.4.1) \qquad W = \sum_{j=1}^{\nu} \underline{y}_j \underline{y}_j',$$

and we shall now concern ourselves with the joint density of the $\frac{1}{2}m(m+1)$ variables: $w_{11}, w_{12}, \ldots, w_{1m}$; $w_{22}, w_{23}, \ldots, w_{2m}$; $w_{33}, w_{34}, \ldots, w_{3m}; \ldots; w_{m-1m-1} w_{m-1m}$; w_{mm}. These $\frac{1}{2}m(m+1)$ variables will have a joint density in $\mathscr{S}_{\frac{1}{2}m(m+1)}$ if and only if $\nu \geq m$; for is $\nu < m$, W is singular so that there exists at least one non-trivial set of constants $\ell_1, \ell_2, \ldots, \ell_m$ such that

$$\sum_{j=1}^{m} \ell_j w_{ij} = 0, \quad \text{all } i = 1, 2, \ldots, m.$$

It may be noted that if $\nu \geq m$, then W is positive definite (positive semidefinite if $\nu < m$), and if $\underline{y}_j$ may be any point in $\mathscr{S}_m$, then W may be any positive definite matrix.

We shall be concerned in this section with the case $Y = (\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_\nu) \frown N_{m \times \nu}((0); V \otimes I)$ so that the $(\underline{y}_j)_1^\nu$ are a set of independent m-variate normal vectors with mean vectors $\underline{0}$ and common dispersion matrix V and will obtain the joint density of the $\frac{1}{2}m(m+1)$ linearly independent elements w_{ij} $(i \leq j)$ of W (it is convenient to write f(W) for the density of these $\frac{1}{2}m(m+1)$ elements provided that it be remembered what is meant by this device).

By a theorem in the algebra of matrices we may write

$$(3.4.2) \qquad W = YY' = TT'$$

where T is lower triangular since W is symmetric; moreover, since W is positive definite, the diagonal elements of T are strictly positive. We suppose T to be of the form

(3.4.3)
$$T = \begin{bmatrix} t_{11} & 0 & 0 & 0 & \cdots & 0 \\ t_{21} & t_{22} & 0 & 0 & \cdots & 0 \\ t_{31} & t_{32} & t_{33} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ t_{m1} & t_{m2} & t_{m3} & t_{m4} & \cdots & t_{mm} \end{bmatrix}$$

with $t_{ii} > 0$ $(i = 1, 2, \ldots, m)$ and note that

(3.4.4)
$$w_{\alpha\beta} = \sum_{j=1}^{\nu} y_{\alpha j} y_{\beta j} = \sum_{i=1}^{\beta} t_{\alpha i} t_{\beta i} \qquad (\alpha \geq \beta).$$

Proceeding via m rather similar stages, we find the joint density of the $\frac{1}{2}m(m+1)$ non-zero elements of T and hence, using (3.4.4) to the joint density of $(w_{\alpha\beta})_{\alpha \geq \beta}$. It will be helpful if the reader familiarizes himself with the polar transformation and in particular remembers that the surface area of a sphere of radius r in p-dimensions is

$$S_p r^{p-1} = 2\pi^{\frac{1}{2}p} r^{p-1} \Big/ \Gamma(\tfrac{1}{2}p)$$

with S_p the surface area of the p-dimensional unit sphere.

The joint density of the elements of Y is

(3.4.5)
$$f\left\{\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1\nu} \\ \vdots & \vdots & & \vdots \\ y_{m1} & y_{m2} & & y_{m\nu} \end{bmatrix}\right\} = \frac{|V|^{-\frac{1}{2}\nu}}{(2\pi)^{\frac{1}{2}m\nu}} \exp\Big(-\frac{1}{2}\sum_{j=1}^{\nu} \underline{y}_j' V^{-1} \underline{y}_j\Big)$$

$$= \frac{|V|^{-\frac{1}{2}\nu}}{(2\pi)^{\frac{1}{2}m\nu}} \exp\Big(-\frac{1}{2}\operatorname{tr} V^{-1}\big(\sum_{j=1}^{\nu} \underline{y}_j \underline{y}_j'\big)\Big).$$

For convenience we write

$$(3.4.6)\quad f\left((y_{ij})_{1,1}^{m,v}\right) = K\,\Phi\left\{\begin{bmatrix} \sum y_{1j}^2 & \sum y_{1j}y_{2j} & \cdots & \sum y_{1j}y_{mj} \\ \vdots & & & \\ \sum y_{mj}y_{1j} & \sum y_{mj}y_{2j} & \cdots & \sum y_{mj}^2 \end{bmatrix}\right\}$$

where K is a constant $(=|V|^{-\frac{1}{2}v}(2\pi)^{-\frac{1}{2}vm})$ and where $\Phi(C)$ is a function of the elements of a matrix C; in our case $\Phi(C) = \exp -\frac{1}{2}\,\mathrm{tr}V^{-1}C$. The summations in (3.4.6) are over j going from 1 to v.

Stage 1. The variables $y_{11}, y_{12}, \ldots, y_{1v}$ are held fixed and an orthogonal matrix (Q, say) is defined such that its first column is

$$\frac{1}{\sqrt{\sum y_{1j}^2}}\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1v} \end{bmatrix}$$

we now define z_{ij}, $i = 1, 2, 3, \ldots, m$; $j = 1, 2, \ldots, v$ by

$$(3.4.7)\qquad (z_{i1}, z_{i2}, \ldots, z_{iv}) = (y_{i1}, y_{i2}, \ldots, y_{iv})Q$$

so that

$$(3.4.8)\qquad \sum_{j=1}^{v} z_{\alpha j}z_{\beta j} = \sum_{j=1}^{v} y_{\alpha j}y_{\beta j} \qquad \substack{\alpha\\ \beta} = 2, 3, \ldots, m$$

and

$$(3.4.9)\qquad z_{\alpha 1}\left(\sum_{j=1}^{v} y_{1j}^2\right)^{1/2} = \sum_{j=1}^{v} y_{\alpha j}y_{1j} \quad \alpha = 2, 3, \ldots, m.$$

[Comparing (3.4.9) and (3.4.4) with $\beta=1$, we see that $\sum y_{1j}^2 = t_{11}^2$, $\sum y_{\alpha j}y_{1j} = t_{\alpha 1}t_{11}$ so that $z_{\alpha 1}$ is the same as $t_{\alpha 1}$.]

The Jacobian of the transformation $(y_{ij})_{2,1}^{m,v} \Rightarrow \{z_{ij}\}_{2,1}^{m,v}$ is unity; we have therefore

(3.4.10) $$f\left\{(y_{1j})_{j=1}^{v};\ (z_{ij})_{2,1}^{m,v}\right\} =$$

$$K\,\Phi\left\{\left(\left[\begin{array}{c:cccc} \sum y_{1j}^2 & z_{21}(\sum y_{1j}^2)^{1/2} & \cdot & \cdot \cdots & z_{m1}(\sum y_{1j}^2)^{1/2} \\ \hdashline z_{21}(\sum y_{1j}^2)^{1/2} & \sum z_{2j}^2 & \sum z_{2j}z_{3j} & \cdots & \sum z_{2j}z_{mj} \\ \vdots & \vdots & & & \\ z_{m1}(\sum y_{1j}^2)^{1/2} & \sum z_{mj}z_{2j} & \cdots & \cdots & \sum z_{mj}^2 \end{array}\right]\right)\right\}$$

Now make a polar transformation on the variables $(y_{ij})_1^v$ of radius $t_{11} = (\sum y_{1j}^2)^{1/2}$, integrate out the polar angles (so that the constant S_v is obtained) and replace z_{a1} by t_{a1} $(a=2,3,\ldots,m)$ to give

(3.4.11) $$f\left\{(t_{a1})_{a=1}^{m}\ (z_{ij})_{2,2}^{m,v}\right\} = KS_v t_{11}^{v-1} \quad X$$

$$\Phi\left\{\left[\begin{array}{c:cccc} t_{11}^2 & t_{21}t_{11} & t_{31}t_{11} & \cdots & t_{m1}t_{11} \\ \hdashline t_{21}t_{11} & \sum z_{2j}^2+t_{21}^2 & \cdot & \cdots & \vdots \\ \vdots & & & & \vdots \\ t_{m1}t_{11} & \sum z_{mj}z_{1j}+t_{m1}t_{11} & \cdots & \cdots & \sum z_{mj}^2+t_{m1}^2 \end{array}\right]\right\}$$

where the summation is the lower right partition is over j going from 2 to v and the general term is

$$\sum_{j=2}^{v} z_{\alpha j} z_{\beta j} + t_{\alpha 1} t_{\beta 1}$$

Stage 2. The processes of Stage 1 are repeated on the variables $(z_{ij})_{2,2}^{m,v}$. Define an orthogonal matrix (R, say) such that its first column is

$$\frac{1}{\sqrt{\sum_{j=2}^{v} z_{2j}^2}} (z_{22}, z_{23}, \ldots, z_{2v})'$$

and define u_{ij}, $i = 3, 4, \ldots, m$; $j = 2, 3, \ldots, v$ by

$$(u_{i2}, u_{i3}, \ldots, u_{iv}) = (z_{i2}, z_{i3}, \ldots, z_{iv})R \tag{3.4.12}$$

so that

$$\begin{aligned} \sum_{j=2}^{v} u_{\alpha j} u_{\beta j} &= \sum_{j=2}^{v} z_{\alpha j} z_{\beta j} && \text{(by 3.4.12)} \quad {}^{\alpha}_{\beta}=2,3,\ldots,m \\ &= -t_{\alpha 1} t_{\beta 1} + \sum_{j=1}^{v} y_{\alpha j} y_{\beta j} && \text{(by 3.4.8)} \quad {}^{\alpha}_{\beta}=2,3,\ldots,m \\ &= \sum_{i=2}^{\beta} t_{\alpha i} t_{\beta i} && \text{(by 3.4.4)} \quad {}^{\alpha}_{\beta}=2,3,\ldots,m. \end{aligned} \tag{3.4.13}$$

Also

$$\begin{aligned} u_{\alpha 1}\Big(\sum_{j=2}^{v} z_{2j}^2\Big)^{\frac{1}{2}} &= \sum_{j=2}^{v} z_{\alpha j} z_{2j} && \text{(by 3.4.11 and 12)} \quad \alpha=3,4,\ldots,m \\ &= -t_{\alpha 1} t_{21} + \sum_{j=1}^{v} y_{\alpha j} y_{2j} && \text{(by 3.4.8)} \quad \alpha=3,4,\ldots,m \\ &= t_{\alpha 2} t_{22} && \text{(by 3.4.4)} \quad \alpha=3,4,\ldots,m. \end{aligned} \tag{3.4.14}$$

[From the equation (3.4.14), we see that

$$\sum_{j=2}^{v} z_{2j}^2 = t_{22}^2 \; ; \quad \sum_{j=2}^{v} z_{\alpha j} z_{2j} = t_{\alpha 2} t_{22}$$

so that $u_{\alpha 2}$ is the same as $t_{\alpha 2}$.]

The Jacobian of the transformation $(z_{ij})_{3,2}^{m,v} \to (u_{ij})_{3,2}^{m,v}$ is unity; we have therefore,

(3.4.15)
$$f\left\{(t_{i1})_{i=1}^{m} ; (z_{2j})_{j=2}^{v} ; (u_{ij})_{3,2}^{m,v}\right. =$$

$$K \,\Phi \left\{ \left[\begin{array}{cc|c} t_{11}^2 & t_{21}t_{11} & t_{31}t_{11} \cdot \cdot \cdot \cdot \cdot \, t_{m1}t_{11} \\ t_{21}t_{11} & t_{21}^2 + u_{22}(\sum z_{2j}^2)^{\frac{1}{2}} & \cdot\cdot\cdot\cdot\cdot \quad t_{m1}t_{21} + u_{m2}(\sum z_{2j}^2)^{\frac{1}{2}} \\ \hline & & \sum_{j=2}^{v} u_{\alpha j} u_{\beta j} + t_{\alpha 1} t_{\beta 1} \\ t_{m1}t_{11} & t_{m1}t_{21} + u_{m2}(\sum z_{2j}^2)^{\frac{1}{2}} & \end{array} \right] \right\}$$

all summations being over j going from 2 to v. Now make a polar transformation on the variables $(z_{2j})_{j\,2}^{v}$ of radius $(\sum_{j=2}^{v} z_{2j}^2)^{1/2} = t_{22}$. Integrate out the polar angles (so that the constant S_{v-1} is obtained) and replace $u_{\alpha 2}$ by $t_{\alpha 2}$ $(\alpha = 3, 4, \ldots, m)$ to give:

(3.4.16)
$$f\left\{(t_{\alpha 1})_{\alpha=1}^{m} ; (t_{\alpha 2})_{\alpha=2}^{m} ; (u_{ij})_{3,3}^{m,v}\right\} = K S_v S_{v-1} t_{11}^{v-1} t_{22}^{v-2} \; X$$

$$\Phi \left\{ \left[\begin{array}{cc|c} t_{11}^2 & t_{21}t_{11} & t_{31}t_{11} \cdot \cdot \cdot \cdot \cdot \, t_{m1}t_{11} \\ t_{21}t_{11} & t_{21}^2 + t_{22}^2 & t_{31}t_{21} + t_{32}t_{22} \cdot \cdot \cdot \, t_{m1}t_{21} + t_{m2}t_{22} \\ \hline \vdots & & \\ t_{m1}t_{11} & & t_{\alpha 1}t_{\beta 1} + t_{\alpha 2}t_{\beta 2} + \sum_{j=3}^{v} z_{\alpha j} z_{\beta j} \end{array} \right] \right\}$$

Proceeding in this manner thru stage m, we have finally, since $\nu \geq m$,

$$(3.4.17)\qquad f\left\{(t_{ij})_{i=1;j=1}^{i=m;j=m}\right\} = KS_\nu S_{\nu-1}\cdots S_{\nu-m+1}t_{11}^{\nu-1}t_{22}^{\nu-2}\cdots t_{m-1;m-1}^{\nu-m+1}t_{m;m}^{\nu-m}$$

$$\times\ \Phi\left\{\sum_{i=1}^{\min(\alpha,\beta)} t_{\alpha i}t_{\beta i}\right\},$$

that is,

$$(3.4.18)\qquad f(T) = KS_\nu S_{\nu-1}\cdots S_{\nu-m+1}t_{11}^{\nu-1}\cdots t_{mm}^{\nu-m}\ \Phi(TT').$$

We now transform to $(w_{\alpha\beta})_{m\geq\alpha\geq\beta\geq 1}$ using the equations (3.4.4); these are $w_{\alpha\beta} = \sum_{j=1}^{\beta} t_{\alpha j}t_{\beta j}$. The Jacobian,

$$J = \left\|\frac{\partial(w_{11},w_{21},\ldots,w_{m1};\ w_{22},w_{32},\ldots,w_{m2};\ldots;w_{mm})}{\partial(t_{11},t_{21},\ldots,t_{m1};\ t_{22},t_{32},\ldots,t_{m2};\ldots;t_{mm})}\right\|$$

is triangular since $\frac{\partial w_{\alpha\beta}}{dt_{ij}} = 0$ whenever $i>\alpha$. Hence

$$(3.4.19)\qquad J = \prod_{\alpha=1}^{m}\prod_{\beta=1}^{\alpha}\frac{\partial w_{\alpha\beta}}{\partial t_{\alpha\beta}}$$

But

$$\frac{dw_{\alpha\beta}}{dt_{\alpha\beta}} = \begin{cases}2t_{\alpha\alpha} & \alpha=\beta\\ t_{\beta\beta} & \alpha>\beta.\end{cases}$$

Hence

(3.4.20) $$J = 2^m t_{11}^m t_{22}^{m-1} \ldots . t_{m-1:m-1}^1 t_{m:m}^0$$

so that

(3.4.21) $$f(W) = 2^{-m} K S_\nu S_{\nu-1} \ldots S_{\nu-m+1} \Phi(W) . \prod_{i=1}^{m} t_{ii}^{\nu-m-1}.$$

Since $|T| = t_{11}t_{22}\ldots t_{mm} = |W|^{\frac{1}{2}}$ and simplifying the constant term, we have finally

*(3.4.22) $$f(W) = \frac{|V|^{-\frac{1}{2}\nu}}{2^{\frac{1}{2}\nu m} \pi^{\frac{1}{4}m(m-1)}} . \frac{1}{\prod_{j=1}^{m} \Gamma(\frac{\nu+1-j}{2})} . |W|^{\frac{\nu-m-1}{2}} e^{-\frac{1}{2}\mathrm{tr}V^{-1}W}.$$

Observe that when $m = 1$, so that W is scalar ($=w$, say) then, with $V = \sigma^2$, scalar,

(3.4.23) $$f(w) = \frac{w^{\frac{1}{2}\nu-1} e^{\frac{-w}{2\sigma^2}}}{2^{\frac{1}{2}\nu}\sigma^\nu \Gamma(\frac{1}{2}\nu)}$$

which is the χ^2-density with ν degrees of freedom.

It will be convenient to write the constant part of (3.4.22) as

---

*This is also a valid density when ν is not integer ($\nu > m-1$) see Section 3.8. At this point, however, we have only established $f(W)$ as a density for ν an integer ($\geq m$).

$K_m(V,\nu)$ and write

$$f(W) = K_m(V,\nu)|W|^{\frac{\nu-m-1}{2}}e^{-\frac{1}{2}\mathrm{tr}V^{-1}W} \tag{3.4.24}$$

with

$$K_m(V,\nu) = 1 \Big/ 2^{\frac{1}{2}\nu m}\pi^{\frac{1}{4}m(m-1)}|V|^{\frac{1}{2}\nu}\prod_{j=1}^{m}\Gamma\left(\frac{\nu+1-j}{2}\right) \tag{3.4.25}$$

3.5 The distribution of the trace of a Wishart variate.

We suppose W to be a Wishart variate

$$W \sim \mathcal{W}_m(V;\nu:\Lambda) \qquad \nu = 1, 2,\ldots, m,\ldots \tag{3.5.1}$$

and obtain the distribution of trW.

There exists orthogonal P such that $PVP' = D_V$ with D_V a diagonal matrix of latent roots of V.

Now let $W_0 = PWP'$ and by Theorem 15

$$W_0 \sim \mathcal{W}_m(D_V;\nu:P\Lambda P'); \tag{3.5.2}$$

so that

$$\mathrm{tr}W = \mathrm{tr}\,W_0 \sim \sum_{i=1}^{m}(D_V)_{ii}\cdot{}_{(i)}\chi^2_\nu\left(\frac{1}{2}\frac{(P\Lambda P')_{ii}}{(D_V)_{ii}}\right) \tag{3.5.3}$$

where $\{{}_{(i)}\chi^2_\nu\}_{i=1}^{m}$ are a set of independent (since $\{(W_0)_{ii}\}_{i=1}^{m}$ are independent) χ^2_ν-variates with noncentrality parameters as shown.

An important consequence of the result expressed in equation

(3.5.3) is:

Theorem 17. (See also Theorer 27, Section 3.8).

If $W \sim \mathcal{W}_m(\sigma^2 I; \nu; \Lambda)$, $\nu = 1, 2, \ldots, m, \ldots$; then with $\lambda = \frac{1}{2\sigma^2}\operatorname{tr}\Lambda$ and $\eta = \operatorname{tr}W$

$$f(\eta) = e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \frac{\eta^{\frac{1}{2}m\nu+j-1} e^{-\eta/2\sigma^2}}{(2\sigma^2)^{\frac{1}{2}m\nu+j}\,\Gamma(\frac{1}{2}m\nu+j)} ;$$

that is

$$\eta \sim \sigma^2 \chi^2_{m\nu}\left(\frac{1}{2\sigma^2}\operatorname{tr}\Lambda\right)$$

3.6 Functions of the Wishart variate having a Wishart distribution.

We have seen in the last section that the trace of a Wishart variate has a χ^2-distribution (which is a Wishart variate of dimension one) when the dispersion matrix is $\sigma^2 I$. There are two functions of the Wishart variate which enter into the theory frequently and which themselves have a Wishart distribution.

Suppose W to be partitioned as

$$W = \left(\begin{array}{c|c} W_{11} & W_{12} \\ \hline W_{21} & W_{22} \end{array}\right) \tag{3.6.1}$$

with W_{11} a pxp symmetric and positive definite matrix and W_{22} a qxq symmetric positive definite matrix. By the symmetry of W, $W_{12} = W_{21}'$ is pxq.

Theorem 18. (See also Theorem 28, Section 3.8).

If $W \sim \mathcal{W}_m(V; v; \Lambda)$, $v = 1, 2, \ldots, m, \ldots$; then $W_{11} \sim \mathcal{W}_p(V_{11}; v; \Lambda_{11})$ where V_{11} and Λ_{11} are corresponding partitions of V and Λ, and p is the dimension of W_{11}.

The proof is obvious. Let $W = ZZ'$ where $Z \frown N_{m\times v}(\xi; V \otimes I)$ such that $\xi\xi' = \Lambda$ so that W has the required distribution. Partition Z by writing

$$Z = \begin{bmatrix} Z_{(1)} \\ \cdots \\ Z_{(2)} \end{bmatrix} \tag{3.6.2}$$

with $Z_{(1)}$ a $p\times v$ and $Z_{(2)}$ a $q\times v$ matrix, then

$$\begin{cases} Z_{(1)} \frown N_{p\times v}(\xi_{(1)}; V_{11} \otimes I) \\ \\ Z_{(2)} \frown N_{q\times v}(\xi_{(2)}; V_{22} \otimes I), \end{cases} \tag{3.6.3}$$

with $\xi_{(1)}$, $\xi_{(2)}$, partitions corresponding to the partitions of Z and

$$V = \left(\begin{array}{c|c} V_{11} & V_{12} \\ \hline V_{21} & V_{22} \end{array}\right)$$

a corresponding partition of V.

Clearly

$$W_{11} = Z_{(1)}Z'_{(1)} \tag{3.6.4}$$

and by definition has the distribution $\mathcal{W}_p(V_{11};v:\Lambda_{11})$ since $\Lambda_{11} = \xi_{(1)}\xi'_{(1)}$.

It may be noted also that $W_{22} \sim \mathcal{W}_q(V_{22};v:\Lambda_{22})$ and that W_{11} and W_{22} have independent distributions if $V_{12} = (0)$.

We consider now the (qxq) matrix $W_{22}-W_{21}W_{11}^{-1}W_{12}$. This is a symmetric matrix defined whenever W_{11}^{-1} exists, that is whenever $p \leq v$. In terms of the partitions of Z defined in (3.6.2), we may write

$$\begin{aligned} W_{22}-W_{21}W_{11}^{-1}W_{12} &= Z_{(2)}Z'_{(2)}-Z_{(2)}Z'_{(1)}(Z_{(1)}Z'_{(1)})^{-1}Z_{(1)}Z'_{(2)} \\ &= Z_{(2)}\left(I-Z'_{(1)}(Z_{(1)}Z'_{(1)})^{-1}Z_{(1)}\right)Z'_{(2)}. \end{aligned} \tag{3.6.5}$$

If we hold $Z_{(1)}$ fixed, then we have an idempotent generalised quadratic form in $Z_{(2)}$. The rank of the idempotent matrix is v-p. Hence by Section 3.2:

$$(W_{22}-W_{21}W_{11}^{-1}W_{12})\,\Big|\,Z_{(1)} \sim \mathcal{W}_q(V_{22}-V_{21}V_{11}^{-1}V_{12};v-p:L) \tag{3.6.6}$$

where the noncentrality matrix, L, is given by

$$\begin{aligned} L &= \left(\xi_{(2)}+V_{21}V_{11}^{-1}(Z_{(1)}-\xi_{(1)})\right)\left(I-Z'_{(1)}(Z_{(1)}Z'_{(1)})^{-1}Z_{(1)}\right) \quad \times \\ &\qquad \left(\xi_{(2)}+V_{21}V_{11}^{-1}(Z_{(1)}-\xi_{(1)})\right)' \\ &= (\xi_{(2)}-V_{21}V_{11}^{-1}\xi_{(1)})(I-Z'_{(1)}(Z_{(1)}Z'_{(1)})^{-1}Z_{(1)})(\xi_{(2)}-V_{21}V_{11}^{-1}\xi_{(1)})' \end{aligned} \tag{3.6.7}$$

The conditional distribution of $W_{22}-W_{21}W_{11}^{-1}W_{12}$ given $Z_{(1)}$ is a Wishart variate with noncentrality matrix which is a function of the $Z_{(1)}$ unless

(a) $\underline{\mu} = (0)$, or

(b) $\underline{\mu}_{(2)} = (0)$ and $V_{12} = V'_{21} = (0)$.*

If either (a) or (b) holds, then the conditional distribution of $W_{22}-W_{21}W_{11}^{-1}W_{12}$ is mathematically independent of $Z_{(1)}$ (and therefore W_{11}) so that we get the theorems:

Theorem 19. (See also Theorem 29, Section 3.8).

If $W \sim \mathcal{W}_m(V;\nu(0))$ and then $(W_{22}-W_{21}W_{11}^{-1}W_{12}) \sim \mathcal{W}_q(V_{22}-V_{21}V_{11}^{-1}V_{12};\nu-p:(0))$ and is statistically independent of W_{11}.

This follows from (3.6.6) and (3.6.7) showing that L = (0).

Theorem 20.

If $W \sim \mathcal{W}_m(V;\nu(\Lambda))$, integer with $\nu \geq p$; and if V is of the form $\left(\begin{array}{c|c} V_{11} & (0) \\ \hline (0) & V_{22} \end{array}\right)$ and Λ is of the form $\left(\begin{array}{c|c} \Lambda_{11} & (0) \\ \hline (0) & (0) \end{array}\right)$; V_{11} and Λ_{11} are pxp; then $(W_{22}-W_{21}W_{11}^{-1}W_{12}) \sim \mathcal{W}_q(V_{22};\nu-p:(0))$ and is statistically independent of W_{11}.

---

*The case $\underline{\mu}_{(2)} = V_{21}V_{11}^{-1}\underline{\mu}_{(1)} \neq \underline{0}$ also leads to a Wishart distribution on $W_{22}-W_{21}W_{11}^{-1}W_{12}$, but this case must be rare.

The proof is from (3.6.6) with $V_{12} = V'_{12} = (0)$ and (3.6.7) showing $L = (0)$.

An interesting extension to Theorem 19 obtains immediately. Let $_{(r)}W$ be the matrix of the array left by striking out the last (m-r) rows and column of W so that

$$
{}_{(r)}W = \begin{bmatrix} w_{11} & w_{12} & \cdot\,\cdot & w_{1r} \\ w_{21} & w_{22} & \cdot\,\cdot & w_{2r} \\ \vdots & \vdots & & \vdots \\ w_{r1} & w_{r2} & \cdot\,\cdot & w_{rr} \end{bmatrix} \tag{3.6.8}
$$

then

$$
\left\{\begin{aligned}
&|{}_{(1)}W| = w_{11} \\
&|{}_{(2)}W| \big/ |{}_{(1)}W| = w_{22} - w_{21}w_{11}^{-1}w_{12} \\
&|{}_{(3)}W| \big/ |{}_{(2)}W| = w_{33} - (w_{31}w_{32})\,{}_{(2)}W^{-1}(w_{13}w_{12})' \\
&\qquad\qquad \text{etc.} \\
&|{}_{(r)}W| \big/ |{}_{(r-1)}W| = w_{rr} - (w_{r1}, w_{r2}, \ldots, w_{rr-1})\,{}_{(r-1)}W^{-1} \\
&\qquad\qquad\qquad\qquad \times \begin{bmatrix} w_{1r} \\ w_{2r} \\ \vdots \\ w_{r-1r} \end{bmatrix}
\end{aligned}\right. \tag{3.6.9}
$$

Now, if $W \sim \mathcal{W}_m(V;\nu:(0))$, then the distribution of $|_{(r)}W|/|_{(r-1)}W|$ is independent of $_{(r-1)}W$ and therefore of $|_{(r-1)}W|/|_{(r-2)}W|$ etcetera. In addition $|_{(r)}W|/|_{(r-1)}W| \sim \mathcal{W}_1\left(\frac{|_{(r)}V|}{|_{(r-1)}V|};\nu-r+1:(0)\right)$ and since the scalar Wishart is a χ^2-variate, we have

Theorem 21. (See also Theorem 30, Section 3.8).

If $W \sim \mathcal{W}_m(V;\nu:(0))$, ν integer with $\nu \geq m$; and $_{(r)}W$ is as defined by (3.6.8), then

$$\{w_{11};\{|_{(r)}W| \big/ |_{(r-1)}W|\}_{r=2}^{m}\}$$

are independent χ^2-variates such that

$$w_{11} \sim V_{11}\chi^2_\nu.$$

$$\frac{|_{(r)}W|}{|_{(r-1)}W|} \sim \frac{|_{(r)}V|}{|_{(r-1)}V|}\chi^2_{\nu-r+1}$$

As a final consideration in this section, we consider the partition of (3.6.1) that is

$$W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} = \begin{bmatrix} Z_{(1)} \\ Z_{(2)} \end{bmatrix} \begin{bmatrix} Z'_{(1)} \vdots Z'_{(2)} \end{bmatrix}$$

(W_{11} is pxp; W_{22} is qxq; p+q = m)

with $Z \frown N_{m\times\nu}\big((0);\ V \otimes I\big)$, and consider $W_{22}-W_{21}W_{11}^{-1}W_{12}$ and $W_{21}W_{11}^{-1}W_{12}$. Now:

$$W_{22}-W_{21}W_{11}^{-1}W_{12} = Z_{(2)}\left\langle I-Z'_{(1)}(Z_{(1)}Z'_{(1)})^{-1}Z_{(1)}\right\rangle Z'_{(2)} \tag{3.6.10}$$

and

$$W_{21}W_{11}^{-1}W_{12} = Z_{(2)}\left\langle Z'_{(1)}(Z_{(1)}Z'_{(1)})^{-1}Z_{(1)}\right\rangle Z'_{(2)} \tag{3.6.11}$$

so that conditional on $Z_{(1)}$, $W_{22}-W_{21}W_{11}^{-1}W_{12}$ and $W_{21}W_{11}^{-1}W_{12}$ have independent Wishart distributions; in fact

$$(3.6.12) \qquad (W_{22}-W_{21}W_{11}^{-1}W_{12})|Z_{(1)} \sim \mathcal{W}_q\left(V_{22}-V_{21}V_{11}^{-1}V_{12};\nu-p:(0)\right)$$

(see Theorem 21) and by Section (3.3)

$$(3.6.13) \qquad (W_{21}W_{11}^{-1}W_{12})|Z_{(1)} \sim \mathcal{W}_q\left(V_{22}-V_{21}V_{11}^{-1}V_{12};p:V_{21}V_{11}^{-1}Z_{(1)}Z'_{(1)}V_{11}^{-1}V_{12}\right)$$

and again by Section (3.3) $W_{22}-W_{21}W_{11}^{-1}W_{12}$ and $W_{21}W_{11}^{-1}W_{12}$ have independent distributions conditional on $Z_{(1)}$. Now since the distribution of $W_{22}-W_{21}W_{11}^{-1}W_{12}$ is functionally independent of $Z_{(1)}$, we have the $(W_{22}-W_{21}W_{11}^{-1}W_{12})$ is distributed independently of $W_{21}W_{11}^{-1}W_{12}$; however, the non-conditional distribution of the latter is no longer of the Wishart type. Suppose that $V_{12} = V'_{21} = (0)$, then $W_{21}W_{11}^{-1}W_{12}$ is now functionally independent of $Z_{(1)}$; we have, therefore, the following theorems:

Theorem 22. (See also Theorem 31, Section 3.8).

If $W \sim \mathcal{W}_m(V;\nu:(0))$ and V are similarly partitioned where ν is integer and $\nu \geq p$ then $W_{22}-W_{21}W_{11}^{-1}W_{12}$ and $W_{21}W_{11}^{-1}W_{12}$ are independently distributed such that

$$(W_{22}-W_{21}W_{11}^{-1}W_{12}) \sim \mathcal{W}_q(V_{22}-V_{21}V_{11}^{-1}V_{12};\nu-p:(0)).$$

Theorem 23. (See also Theorem 32, Section 3.8).

If $W \sim \mathcal{W}_m(V;\nu:(0))$ and V are similarly partitioned and if $V_{12} = V'_{21} = (0)$, then $(W_{22}-W_{21}W_{11}^{-1}W_{12})$ and $W_{21}W_{11}^{-1}W_{12}$ have independent distributions such that

$$(W_{22}-W_{21}W_{11}^{-1}W_{12}) \sim \mathcal{W}_q(V_{22};\nu-p:(0)) \text{ and}$$

$$(W_{21}W_{11}^{-1}W_{12}) \sim \mathcal{W}_q(V_{22};p:(0)),$$

whenever W_{11}^{-1} exists; that is for $\nu \geq p$.

3.7 The distribution of the determinant of a Wishart variate.

Very little can be done for the general noncentral Wishart variate. We look at the central case only therefore.

Theorem 24. (See also Theorem 33, Section 3.8).

If $W \sim \mathcal{W}_m(V;\nu:(0))$; ν integer with $\nu \geq m$; $|W| \sim |V| \prod_{j=1}^{m} \chi^2_{\nu+j-j}$.

We may prove this theorem after invoking Theorem 21. Since

$$|W| = w_{11} \prod_{j=2}^{m} \frac{|_{(j)}W|}{|_{(j-1)}W|} \tag{3.7.1}$$

and using the statistical independence of the set

$$\{w_{11}\ \{|_{(j)}W|/|_{(j-1)}W|\}_2^m\}$$

we have

$$|W| \sim v_{11} \prod_{j=2}^{m} \frac{|_{(j)}V|}{|_{(j-1)}V|} \prod_{j=1}^{m} \chi^2_{\nu+j-j}(0)$$

that is

$$|W| \sim |V| \prod_{j=1}^{m} \chi^2_{\nu+1-j}(0).$$

An alternative method of proof is to obtain the general moment of $|W|$. We have

$$(3.7.2) \qquad \mathcal{E}|W|^h = \int \cdots \int_{\text{pos.def.}W} K_m(V;\nu)\, |W|^{\frac{1}{2}(\nu-m-1)+h} e^{-\frac{1}{2}\operatorname{tr}V^{-1}W}$$

$$= \frac{K_m(V;\nu)}{K_m(V;\nu+2h)} \int \cdots \int_{\text{pos.def.}W} K_m(V;\nu+2h)\, |W|^{\frac{1}{2}(\nu+2h-m-1)} e^{-\frac{1}{2}\operatorname{tr}V^{-1}W}$$

The integrand is density $\mathcal{W}_m(V;\nu+2h:(0))$ and the integral is consequently unity so that

$$(3.7.3) \qquad \mathcal{E}|W|^h = K_m(V;\nu)\big/K_m(V;\nu+2h)$$

and after simplification

$$*(3.7.4) \qquad \mathcal{E}|W|^h = |V|^h \prod_{j=1}^{m} \left\{ \frac{\Gamma\left(\frac{\nu+1-j}{2}+h\right)}{\Gamma\left(\frac{\nu+1-j}{2}\right)}\, 2^h \right\}$$

which is the h-th moment of the variate which is distributed as

---

*At this point we should require that h and ν are integer with $\nu+2h \geq m$, however, see also Theorem 33.

$$|V| \prod_{j=1}^{m} \chi^2_{\nu+j-j}$$

3.8 The Wishart density (non-integer degrees of freedom).

Looking back over this chapter, we defined the Wishart distribution in terms of a number (ν) of normally distributed vectors; as a consequence, it is implicit in a discussion of the Wishart distribution that the degrees of freedom are integer. In Section 3.4 it was shown that starting from the density of $X \frown N_{m\times\nu}\big((0);\ V \otimes I\big)$ with X an m×ν matrix, we could arrive at the density of $W = XX'$ when $\nu \geq m$. The derivation of this density assumes that ν is integer (since it is one of the dimensions of matrix X). The form of this density is given in equation (3.4.22).

In all practical cases dealing with the problem of estimation or testing given samples from a normal population, any Wishart variates which emerge will have integer degrees of freedom. This statement is to be compared with the univariate situation wherein the test statistics are χ^2-variables, or ratios of χ^2-variables with integer degrees of freedom. However, just as

$$p(\chi^2_\nu) = \frac{(\chi^2)^{\frac{1}{2}\nu-1} e^{-\frac{1}{2}\chi^2}}{2^{\frac{1}{2}\nu}\Gamma(\frac{1}{2}\nu)} \qquad (d\chi^2)$$

is a valid density function over $(0,\infty)$ for any $\nu>0$ (integer or not) so also is

$$f(W) = \frac{|V|^{-\frac{1}{2}\nu}|W|^{\frac{\nu-m-1}{2}}\,e^{-\frac{1}{2}\mathrm{tr}V^{-1}W}}{2^{\frac{1}{2}\nu m}\pi^{\frac{1}{4}m(m-1)}\prod_{j=1}^{m}\Gamma(\frac{\nu+1-j}{2})}$$

a valid density function on $\{w_{ij}\}_{j=1;i=1}^{i;\ n}$ over the domain for which W is positive definite for any $\nu>m-1$ (integer or not).

Clearly, f(W) is non-negative; it remains to show that

$$\int\cdots\int_{|W| \text{ is pos. def.}} f(w) = 1$$

The first step is to make the transformation W = TT' where T is non-singular, lower triangular matrix; the transformation will be one-to-one if we insist that the diagonal elements of T are positive. The distribution of T (c.f. equation (3.4.19)) is

$$f(T) = \frac{|V|^{-\frac{1}{2}\nu}t_{11}^{\nu-1}t_{22}^{\nu-2}\cdots t_{mm}^{\nu-m}e^{-\frac{1}{2}\mathrm{tr}T'V^{-1}T}}{2^{\frac{1}{2}m(\nu-m)}\pi^{\frac{1}{4}m(m-1)}\prod_{j=1}^{m}\Gamma(\frac{\nu+1-j}{2})}$$

$$\text{with } 0<t_{ii}<\infty \qquad 1\le i<m$$

$$-\infty<t_{ij}<\infty \qquad 1\le j<i\le m.$$

There exists non-singular, lower triangular L, with diagonal elements positive such that V = LL'. Write $R = L^{-1}T$ to obtain

$$f(R) = \frac{r_{11}^{\nu-1} r_{22}^{\nu-2} \ldots r_{mm}^{\nu-m} e^{-\frac{1}{2}\sum_{i=1}^{m} \sum_{j=1}^{i} r_{ij}^{2}}}{2^{\frac{1}{2}m\nu-m} \pi^{\frac{1}{4}m(m-1)} \prod_{j=1}^{m} \Gamma(\frac{\nu+1-j}{2})}$$

$$\text{with} \quad 0 \leq r_{ii} < \infty \qquad i = 1, 2, \ldots, m$$

$$\infty < r_{ij} < \infty \qquad 1 \leq j < i \leq m,$$

so that the elements of R have independent distribution such that

$$f(r_{ij}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}r_{ij}^{2}} \qquad 1 \leq j < i \leq m$$

$$f(r_{ii}) = \frac{r_{ii}^{\nu-i} e^{-\frac{1}{2}r_{ii}^{2}}}{2^{\frac{1}{2}(\nu-i)-1} \Gamma(\frac{\nu-i+1}{2})} \qquad i = 1, 2, \ldots, m,$$

or, in short-hand notation

$$r_{ij} \frown N_1(0,1) \qquad 1 \leq j < i \leq m$$

$$r_{ii} \frown \chi_{\nu-i+1}. \qquad i = 1, 2, \ldots, m.$$

provided that $\nu-m+1$ 0 so that r_{mm} has a valid distribution.

Hence

$$\int\cdots\int_{\text{pos. def. } W} f(W) = \int\cdots\int_{\substack{\{0\le r_{ii}<\infty\}_{i=1}^{m} \\ \{-\infty<r_{ij}<\infty\}_{1\le j<i\le m}}} f(R) = 1; \quad (\nu>m-1).$$

We state the main results obtained thus far in this section in two theorems:

Theorem 25.

If $f(W) = f(\{w_{ij}\}_{1\le j\le i\le m}) =$

$$\frac{|V|^{-\frac{1}{2}\nu}|W|^{\frac{\nu-m-1}{2}}e^{-\frac{1}{2}\mathrm{tr}V^{-1}W}}{2^{\frac{1}{2}\nu m}\pi^{\frac{1}{4}m(m-1)}\prod_{j=1}^{m}\Gamma(\frac{\nu+1-j}{2})}$$

where $w_{ij} = w_{ji}$

and $W = (w_{ij})$; then $f(W)$ is a valid density function on the $\frac{1}{2}m(m+1)$ variables $\{w_{ij}\}_{1\le j<i\le m}$ over a domain for which W is positive definite for any $\nu>m-1$;

Theorem 26.

If $W \frown \mathcal{W}_m(V;\nu:(0))$ so that $\nu > m-1$ (ν not necessarily integer) and if L is the unique lower triangular matrix with positive diagonal elements such that $V = LL'$; and if T is the unique lower triangular matrix with positive diagonal elements such that $W = TT'$, then the $\frac{1}{2}m(m+1)$ non-zero elements of $R = LT$ are independent. Further,

$$r_{ij} \frown N_1(0,1) \qquad 1 \le j < i \le m$$

$$r_{ii}^2 \frown \chi^2_{\nu-i+1} \qquad i = 1, 2, \ldots, m.$$

It would be possible to go on and define a noncentral Wishart density with non-integer degrees of freedom by merely requiring the the $\{r_{ij}\}$ of Theorem 26 have non-zero means and that the $\{r_{ii}^2\}$ of that theorem have noncentral χ^2-distributions; however, the noncentral Wishart density with non-integer degrees of freedom is of little practical importance and is not really germane to our work.

Theorems 17-24 which were stated in terms of a variate having a Wishart distribution (i.e., $\nu = 1, 2, \ldots, m-1, m, m+1, \ldots$) are valid for non-integer ν greater than $m-1$. The method of proof is of course necessarily different since for ν non-integer, we cannot represent W as XX' with $X \frown N_{m\times\nu}(\mu;V \otimes I)$.

Instead, we need repeated recourse to Theorem 26. The rest of the theorems are stated with only a sketch of the proof.

Theorem 27. (See also Theorem 17, Section 3.5).

If $W \frown \mathcal{W}_m(\sigma^2 I;\nu:(0))$ with $\nu > m-1$, then

$$\mathrm{tr}W \frown \sigma^2 \chi^2_{m\nu}$$

Theorem 27 follows quickly since $\frac{1}{2}\,\mathrm{tr}W \frown \sum_{i=1}^{m} \sum_{j=1}^{i} r_{ij}^2$

with (r_{ij}) distributed as described in Theorem 26.

Theorem 28. (See also Theorem 18, Section 3.6)

If $W \frown \mathcal{W}_m(V;\nu:(0))$ and W_{11},(pxp), is the leading minor of W, then $W_{11} \frown \mathcal{W}_p(V_{11};\nu:(0))$ where V_{11} is the corresponding minor or V.

In the following theorems we shall partition W as

$$W = \left[\begin{array}{c|c} W_{11} & W_{12} \\ \hline W_{21} & W_{22} \end{array}\right]$$

with

W_{11} a pxp partition $\quad$ $p+q = m$

W_{22} a qxq partition

V is partitioned similarly. The degrees of freedom, ν, are held greater then m-1, but not necessarily integer.

Theorem 29. (See also Theorem 19, Section 3.6).

If $W \sim \mathcal{W}_m(V;\nu:(0))$, any $\nu > m-1$; then, $(W_{22}-W_{21}W_{11}^{-1}W_{12}) \sim \mathcal{W}_q(V_{22}-V_{21}V_{11}^{-1}V_{12};\nu-p:(0))$ and is statistically independent of W_{11}.

To prove Theorem 29, we represent W as TT' with T defined as in the manner of this section. Partition T similarly as for the partition of W, then $W_{11} = T_{11}T_{11}'$

$$W_{22}-W_{21}W_{11}^{-1}W_{12} = T_{22}T_{22}'.$$

But, $T_{11} = L_{11}R_{11}$ and $T_{22} = L_{22}R_{22}$ where $L = \left[\begin{array}{c|c} L_{11} & (0) \\ \hline L_{21} & L_{22} \end{array}\right]$ is the lower triangular matrix such that V = LL' and R is the matrix of Theorem 26. The independence of W_{11} and $W_{22} = W_{21}W_{11}^{-1}W_{12}$ is a consequence of the independence of R_{11} and R_{22}.

Now

$$W_{11} = T_{11}T_{11}' = L_{11}R_{11}R_{11}'L_{11}' \sim \mathcal{W}_p(L_{11}L_{11}';\nu:(0))$$

$$W_{22}-W_{21}W_{11}^{-1}W_{12} = T_{22}T_{22}' = L_{22}R_{22}R_{22}'L_{22} \sim \mathcal{W}_q(L_{22}L_{22}';\nu-p:(0)).$$

It remains only to identify $L_{11}L_{11}' = V_{11}$ and $L_{22}L_{22}' = V_{22}-V_{21}V_{11}^{-1}V_{12}$ to complete the proof of Theorems 28 and 29.

Theorem 30. (See also Theorem 21, Section 3.6).

If $W \frown \mathcal{W}_m(V;\nu:(0))$, any $\nu>m-1$; then

$$\left\{ w_{11}, \left\{ |{}_{(r)}W| \Big/ |{}_{(r-1)}W| \right\}_{r=2}^{m} \right\}$$

are independent χ^2-variates such that

$$w_{11} \frown v_{11}\chi^2_\nu \; ; \qquad \frac{|{}_{(r)}W|}{|{}_{(r-1)}W|} \sim \frac{|{}_{(r)}V|}{|{}_{(r-1)}V|}\,\chi^2_{\nu-r+1}$$

where

$${}_{(r)}W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1r} \\ w_{21} & w_{22} & \cdots & w_{2r} \\ \vdots & \vdots & & \vdots \\ w_{r1} & w_{r2} & & w_{rr} \end{bmatrix}$$

and ${}_{(r)}V$ is similarly defined.

Theorem 31. (See also Theorem 22, Section 3.6)

If $W \frown \mathcal{W}_m(V;\nu:(0))$, any $\nu>m-1$, then $W_{22}-W_{21}W_{11}^{-1}W_{12}$ and $W_{21}W_{11}^{-1}W_{12}$ have independent distributions. In addition,

$$(W_{22}-W_{21}W_{11}^{-1}W_{12}) \frown \mathcal{W}_q(V_{22}-V_{21}V_{11}^{-1}V_{12};\nu-p:(0)).$$

Theorem 32. (See also Theorem 23, Section 3.6).

If $W \frown \mathcal{W}_m(V;\nu:(0))$, any $\nu>m-1$, then $W_{21}W_{11}^{-1}W_{12}$

$\sim \mathcal{W}_q(V_{22};p:(0))$ whenever $V_{12} = (0)$ where V_{12} is the pxq partition in V corresponding to the partition W_{12} in W.

Theorem 33. (See also Theorem 24, Section 3.7).

If $W \sim \mathcal{W}_m(V;\nu:(0))$, any $\nu > m-1$, then

$$|W| \sim |V| \prod_{j=1}^{m} \chi^2_{\nu+1-j}$$

in particular

$$\mathcal{E}|W|^h = |V|^h \prod_{j=1}^{m} 2^h \frac{\Gamma\left(\frac{\nu+1-j}{2}+h\right)}{\Gamma\left(\frac{\nu+1-j}{2}\right)} \quad \text{for all h satisfying}$$

$\nu+2h > m-1$.

The results of Theorem 33 increase the validity of Theorem 24 and equation (3.7.4) to cover non-integer ν and non-integer h. The proof of Theorem 33 rests in writing

$$|W| = |LL'||RR'| = |V| \prod_{i=1}^{m} r_{ii}^2.$$

CHAPTER 4

THE U-STATISTIC; HOTELLING'S T^2

4.1 Introduction

The U-statistic in the multivariate analysis of variance is the counterpart of the F-ratio in the univariate (scalar) analysis of variance. In the univariate case, we are accustomed to test ratios of the form (hypothesis s.s)/(error s.s) although we could equally well consider (error s.s)/(error s.s + hypothesis s.s). It is the latter ratio which is generalized in the multivariate situation, this generalization leading to the U-statistic.

4.2 Definition of the U-statistic; Moments of central U.

Let $E \sim \mathcal{W}_m(V;\nu:(0))$ and $H \sim \mathcal{W}_m(V;p:\Lambda)$, then

$$U = |E| \big/ |E+H| \tag{4.2.1}$$

is a U-statistic.

The central U-variate is defined in terms of central Wishart variates: let $W_1 \sim \mathcal{W}_m(V;\nu_1:(0))$ and $W_2 \sim \mathcal{W}_m(V;\nu_2:(0))$, then

$$U_{m:\nu_1:\nu_2} \sim \frac{|W_2|}{|W_1+W_2|}. \tag{4.2.2}$$

In our definition, it is necessary that $\nu_2 > m-1$; we insist that ν_1 is integer but allow that $\nu_1 < m$ (so that W_2 has a density whereas W_1 may be claimed to have the Wishart distribution only). The density of W_2 is derived in Section 3.4 and is given in equation

(3.4.22); however, we must write W_1 as the sum of ν_1 matrices of the form $\underline{z}_j\underline{z}_j'$ with $(\underline{z}_1,\underline{z}_2,\ldots,\underline{z}_{\nu_1}) \frown N_{m\times\nu_1}((0),V\otimes I)$. We have then

$$(4.2.3) \qquad f(W_2,\{\underline{z}_j\}_1^{\nu_1}) =$$

$$K_m(V,\nu_2)|W_2|^{\frac{1}{2}(\nu_2-m-1)}e^{-\frac{1}{2}\mathrm{tr}V^{-1}W_2} \cdot \frac{|V|^{-\frac{1}{2}\nu_1}}{(2\pi)^{\frac{1}{2}m\nu_1}}\, e^{-\frac{1}{2}\mathrm{tr}V^{-1}\sum_{j=1}^{\nu_1}\underline{z}_j\underline{z}_j'}$$

so that

$$(4.2.4) \qquad \mathcal{E}\,\frac{|W_2|^h}{|W_1+W_2|^h} =$$

$$K_m(V,\nu_2)\int\!\!\cdots\!\!\int \frac{|W_2|^{\frac{1}{2}(\nu_2+2h-m-1)}e^{-\frac{1}{2}\mathrm{tr}V^{-1}W_2}}{|W_2+\sum_{j=1}^{\nu_1}\underline{z}_j\underline{z}_j'|^h}\;\frac{|V|^{-\frac{1}{2}\nu_1}}{(2\pi)^{\frac{1}{2}m\nu_1}}\, e^{-\frac{1}{2}\mathrm{tr}V^{-1}\sum_{j=1}^{\nu_1}\underline{z}_j\underline{z}_j'}$$

and where the region of integration is $-\infty<z_{ij}<+\infty \quad \begin{cases} i=1,2,\ldots,m \\ j=1,2,\ldots,\nu_1 \end{cases}$

and W_2 positive definite.

$$(4.2.5)\qquad \mathcal{E}\frac{|W_2|^h}{|W_2+W_1|^h} =$$

$$\frac{K_m(V,\nu_2)}{K_m(V,\nu_2+2h)}\int\cdots\int\frac{|V|^{-\frac{1}{2}\nu_2}\, e^{-\frac{1}{2}\mathrm{tr}V^{-1}\sum\underline{z}_j\underline{z}_j'}}{(2\pi)^{\frac{1}{2}m\nu_2}|W+\sum_1^{\nu_1}\underline{z}_j\underline{z}_j'|^h}\, p(W:V,\nu_2+2h)$$

where $p(W:V,\nu_2+2h)$ is the Wishart density on W with ν_2+2h degrees of freedom and dispersion matrix V.

Now

$$(4.2.6)\qquad W+\sum_1^{\nu_1}\underline{z}_j\underline{z}_j' \wedge \mathcal{W}_m(V;\nu_1+\nu_2+2h:(0))$$

Hence integrating out the $\{z_{ij}\}$ subject to $W+\sum_1^{\nu_1}\underline{z}_j\underline{z}_j'=C$ we get

$$(4.2.7)\qquad \mathcal{E}\frac{|W_2|^h}{|W_2+W_1|^h} =$$

$$\frac{K_m(V,\nu_2)}{K_m(V,\nu_2+2h)}\int\cdots\int_{\text{pos.def.C.}}\frac{K_m(V,\nu_1+\nu_2+2h)|C|^{\frac{1}{2}(\nu_1+\nu_2+2h-m-1)}\exp-\frac{1}{2}\mathrm{tr}V^{-1}C}{|C|^h}$$

$$= \frac{K_m(V,\nu_2)}{K_m(V,\nu_2+2h)} \quad \frac{K_m(V,\nu_1+\nu_2+2h)}{K_m(V,\nu_1+\nu_2)} \quad X$$

$$\underset{\text{pos.def.C.}}{\int\cdots\int} K_m(V,\nu_1+\nu_2)|C|^{\frac{1}{2}(\nu_1+\nu_2-m-1)}e^{-\frac{1}{2}\mathrm{tr}V^{-1}C}.$$

This last integral is the complete integral over $p(C:V,\nu_1+\nu_2)$ and is therefore unity. Hence,

$$(4.2.8) \quad \mathscr{E}(U_{m:\nu_1:\nu_2})^h = \frac{K_m(V,\nu_2)K_m(V,\nu_1+\nu_2+2h)}{K_m(V,\nu_2+2h)K_m(V,\nu_1+\nu_2)}$$

and, using equation (3.4.25), after simplification,

$$(4.2.9) \quad \mathscr{E}(U_{m:\nu_1:\nu_2})^h = \prod_{j=1}^{m} \frac{\Gamma(\frac{\nu_2+1-j}{2}+h)\Gamma(\frac{\nu_1+\nu_2+1-j}{2})}{\Gamma(\frac{\nu_2+1-j}{2})\ \Gamma(\frac{\nu_1+\nu_2+1-j}{2}+h)}.$$

Notice that when m=1

$$(4.2.10) \quad \mathscr{E}(U_{1:\nu_1:\nu_2})^h = \frac{\Gamma(\frac{\nu_2}{2}+h)\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_2}{2})\Gamma(\frac{\nu_1+\nu_2}{2}+h)}$$

$$= \frac{B(\frac{\nu_2}{2}+h;\frac{\nu_1}{2})}{B(\frac{\nu_2}{2};\ \frac{\nu_1}{2})}$$

which is the h-th moment of

$$(4.2.11)\quad p(z) = \frac{1}{B(\frac{\nu_1}{2};\frac{\nu_2}{2})}\; z^{\frac{1}{2}\nu_1-1}(1-z)^{\frac{1}{2}\nu_1-1}$$

the Pearson Type I density (or $B(\frac{\nu_2}{2},\frac{\nu_1}{2})$ - density).

Clearly,

$$(4.2.12)\quad U_{1:\nu_1:\nu_2} \sim \chi^2_{\nu_2}/(\chi^2_{\nu_1} + \chi^2_{\nu_2})$$

so that

$$(4.2.13)\quad \frac{1-U_{1:\nu_1:\nu_2}}{U_{1:\nu_1:\nu_2}} \sim \frac{\chi^2_{\nu_1}}{\chi^2_{\nu_2}} \sim \frac{\nu_1}{\nu_2}\, F_{\nu_1:\nu_2}$$

with $F_{\nu_1:\nu_2}$ the Fisher's F-ratio.

We note that (4.2.9) may be written

$$(4.2.14)\quad \mathcal{E}(U_{m:\nu_1:\nu_2})^h = \prod_{j=1}^{m} \frac{B(\frac{\nu_2+1-j}{2}+h;\frac{1}{2}\nu_1)}{B(\frac{\nu_2+1-j}{2};\frac{1}{2}\nu_1)}$$

so that evidently

$$(4.2.15)\quad U_{m:\nu_1:\nu_2} \sim \prod_{j=1}^{m} z_j$$

where the $\{z_j\}_1^m$ are independent B-variates such that

$$(4.2.16)\quad p(z_j) = \frac{1}{B(\frac{1}{2}(\nu_2+1-j);\frac{1}{2}\nu_1)}\, z_j^{\frac{1}{2}(\nu_2+1-j)-1}(1-z_j)^{\frac{1}{2}\nu_1-1}.$$

It is interesting that the moments of "central" U are independent of V although this may have been anticipated since

$$(4.2.17)\quad \frac{|W_2|}{|W_1+W_2|} = \frac{|MW_2M'|}{|MW_1M'+MW_2M'|}$$

for any non-singular M. Since

$$(4.2.18)\quad \begin{aligned} MW_1M' &\sim \mathcal{W}_m(MVM';\nu_1:(0)) \\ MW_2M' &\sim \mathcal{W}_m(MVM';\nu_2:(0)) \end{aligned}$$

and since there exists M such that $MVM'=I$ we would expect that the moments of $|W_2|/|W_1+W_2|$ are independent of V when the W_1 and W_2 are central.

We turn now to the case where W_1 has a non-central Wishart distribution.

4.3 The noncentral U-variate.

If W_1 has a noncentral Wishart distribution, W_2 remaining central, we shall say that the ratio defined in (4.2.2) has a noncentral U-distribution and write $|W_2|/|W_1+W_2| \sim U'_{m:\nu_1:\nu_2}$

Now

$$(4.3.1) \qquad \frac{|W_2|}{|W_2+W_1|} = \frac{|MW_2M'|}{|MW_2M'+MW_1M'|}$$

and since, if V is positive definite and Λ is positive semi-definite, there exists non-singular M such that

$$MVM' = I$$
$$M\Lambda M' = \Phi$$

with Φ a diagonal matrix with elements ϕ_i which are solutions of $|\Lambda-\phi V| = 0$; it follows that the distribution of $|W_2|/|W_2+W_1|$ when $W_1 \sim \mathcal{W}_m(V;\nu_1:\Lambda)$ and $W_2 \sim \mathcal{W}_m(V;\nu_2:(0))$ is as the distribution of $|C_2|/|C_1+C_2|$ where

$$C_1 = MW_1M' \sim \mathcal{W}_m(I;\nu_1:\Phi)$$
$$C_2 = MW_2M' \sim \mathcal{W}_m(I;\nu_2:(0))$$

C_1 and C_2 being independently distributed. The distribution of $|W_2|/|W_1+W_2|$ is a function then of ν_1,ν_2 and of the latent roots of $V^{-1}\Lambda$ where Λ is the noncentrality parameter of W_1. Notice that Λ and therefore $V^{-1}\Lambda$ is of rank no more than $\min(m;\nu_1)$, that is $V^{-1}\Lambda$ has at least $\max(0;m-\nu_1)$ latent roots which are zero.

4.4 The relationship between $U'_{m:\nu_1:\nu_2}$ and $U'_{\nu_1:m:\nu_1+\nu_2-m}$

We have seen in the previous section that the definition of the noncentral U-variate in terms of $W_1 \sim \mathcal{W}_m(V;\nu_1:\Lambda)$ and

$W_2 \sim \mathcal{W}_m(V;\nu_2:(0))$ may be reduced to a definition of the noncentral U-variate in terms of C_1 and C_2 with

$$(4.4.1)\qquad \begin{aligned} &C_1 \sim \mathcal{W}_m(I;\nu_1:\Phi) &&(\nu_1 \text{ integer}) \\ &C_2 \sim \mathcal{W}_m(I;\nu_2:(0)) &&(\nu_2 > m-1) \end{aligned}$$

with Φ a diagonal matrix with at most $\min(\nu_1:m)$ non-zero diagonal elements. Suppose now that $\nu_1 < m$, then we will represent C_1 by

$$(4.4.2)\quad C_1 = XX'$$

where

$$(4.4.3)\qquad X \sim N_{m\times\nu_1}(T;I \otimes I)$$

such that

$$(4.4.4)\qquad TT' = \Phi$$

Then

$$(4.4.5)\qquad U'_{m:\nu_1:\nu_2} \sim \frac{|C_2|}{|C_1+C_2|} = \left|I+X'C_2^{-1}X\right|^{-1}$$

Define a matrix R to be orthogonal and of the form

$$(4.4.6)\qquad R = \underset{m\times(\nu_1 \vdots m-\nu_1)}{(XA \vdots B)}.$$

This is possible since it requires the existance of matrices A, $(\nu_1 x \nu_1)$ and B $(m x \overline{m-\nu_1})$ such that

$$\begin{aligned} A'A'XA &= I & (\nu_1 x \nu_1) \\ A'X'B &= (0) & (\nu_1 x \overline{m-\nu_1}) \\ B'B &= I & (\overline{m-\nu_1} x \overline{m-\nu_1}) \end{aligned}$$

and since $X'X$ is positive definite $(\nu_1 \leq m)$, A exists; and since $A'X'$ is $(\nu_1 x m)$ of rank ν_1, then B exists whenever $\nu_1 < m$. Using the orthogonality of R, we may write

$$(4.4.7) \qquad U'_{m:\nu_1:\nu_2} \sim |I+X'R(R'C_2R)^{-1}R'X|^{-1}$$

$$= |I+(X'XA \,\vdots\, (0))G^{-1}\left(\frac{A'X'X}{(0)}\right)|^{-1}$$

with $(X'XA \,\vdots\, (0))$ a $\nu_1 x(\nu_1 \,\vdots\, m-\nu_1)$ partitioned matrix. Now the condition distribution of $G = R'C_2R$ for fixed X is $\mathcal{W}_m(I,\nu_2:(0))$ and since this is functionally independent of X, the unconditional distribution of G is also $\mathcal{W}_m(I;\nu_2:(0))$.

Partition G as $\left[\begin{array}{c:c} G_{11} & G_{12} \\ \hdashline G_{21} & G_{22} \end{array}\right]$ with G_{11} a $(\nu_1 x \nu_1)$ matrix, then

$$(4.4.8) \qquad (X'XA \,\vdots\, 0)G^{-1}(X'XA \,\vdots\, 0)' =$$

$$X'XA(G_{11}-G_{12}G_{22}^{-1}G_{21})^{-1}A'X'X$$

so that

$$(4.4.9) \qquad U'_{m:\nu_1:\nu_2} \sim |I+X'XA(G_{11}-G_{12}G_{22}^{-1}G_{21})^{-1}A'X'X|^{-1}$$

$$= \frac{|G_{11}-G_{12}G_{22}^{-1}G_{21}|}{|(G_{11}-G_{12}G_{22}^{-1}G_{21}) + A'X'XX'XA|}$$

Now, by Theorem 21,

$$G_{11}-G_{12}G_{22}^{-1}G_{21} \sim \mathcal{W}_{\nu_1}(I;\nu_2-(m-\nu_1):(0))$$

and also

$$A'X'XX'XA = X'X \sim \mathcal{W}_{\nu_1}(I;m:T'T)$$

Finally since $G_{11}-G_{12}G_{22}^{-1}G_{21}$ and $X'X$ have independent distributions and since the non-zero roots of $T'T$ and TT' are the same, we have

Theorem 34.

If W_1 and W_2 have independent distributions such that

$$W_1 \sim \mathcal{W}_m(V;\nu_1:\Lambda) \qquad \text{integer } \nu_1<m$$

$$W_2 \sim \mathcal{W}_m(V;\nu_2:(0)) \qquad \nu_2>m-1.$$

Then

$$\frac{|W_2|}{|W_1+W_2|} \sim \frac{|E_2|}{|E_1+E_2|}$$

where E_1 and E_2 are independently distributed such that

$$E_1 \sim \mathcal{W}\ W_{\nu_1}(I,m:K)$$

$$E_2 \sim \mathcal{W}\ W_{\nu_1}(I,\nu_1+\nu_2-m:(0))$$

where K is a diagonal matrix such that $\{k_{ii}\}_1^{\nu_1}$ contains all the non-zero latent roots of $V^{-1}\Lambda$ and n-{rank($V^{-1}\Lambda$)} zeros.

4.5 Hotelling's T^2

Hotelling's T^2 is the generalization of the square of Student's t; it is also simply related to a U-statistic with somewhat specialized degrees of freedom. Let $\underline{u}$ and W be independent with

$$(4.5.1) \qquad \underline{u} \frown N_m(\underline{\mu}, V)$$

$$(4.5.2) \qquad W \frown \mathcal{W}_m(V;\nu:(0))$$

then the T^2-statistic is defined as $\underline{u}'W^{-1}\underline{u}$, and we will write formally

$$(4.5.3) \qquad \underline{u}'W^{-1}\underline{u} \sim T^2_m(\nu:\tfrac{1}{2}\underline{\mu}'V^{-1}\underline{\mu})$$

where m and ν are parameters of the "degrees of freedom" type and $\lambda=\frac{1}{2}\underline{\mu}'V^{-1}\underline{\mu}$ is the noncentrality parameter.

It is to be noted that when m=1,

$$(4.5.4) \qquad T^2(\nu:\lambda=\frac{\mu^2}{2\sigma^2}) \sim \frac{\chi_1^2(\lambda)}{\chi_1^2(0)} \sim \frac{1}{\nu}F_{1:\nu}(\lambda) \sim \frac{1}{\nu}t_\nu^2(\lambda)$$

so that $T^2_m(\nu:\lambda)$ may be thought of as a generalization of t_ν^2 to m-dimensions.

We may use the results of the previous section to investigate the distribution (and density) of $T_m^2(\nu:\lambda)$. Since

$$(4.5.5) \qquad \left[\begin{array}{c:c} -1 & \underline{u}' \\ \hdashline \underline{u} & W \end{array}\right] = -\,|W + \underline{u}\underline{u}'|$$

$$= -\,|W|(1+\underline{u}'W^{-1}\underline{u})$$

we see that

$$(4.5.6) \qquad \frac{1}{1+\underline{u}'W^{-1}\underline{u}} = \frac{|W|}{|W+\underline{u}\underline{u}'|}$$

But the distribution of W is central Wishart (equation 4.5.2) and

$$(4.5.7) \qquad \underline{u}\ \underline{u}' \sim W_m(V;1\ \underline{\mu}\ \underline{\mu}')$$

Using Theorem 34 with

$$\begin{cases} \nu_1 = 1 \\ \nu_2 = \nu \\ \Lambda = \underline{\mu}\ \underline{\mu}' \quad \text{(mxm, symmetric, of rank 1)} \end{cases}$$

we find that

$$(4.5.7) \qquad \frac{1}{1+\underline{u}'W^{-1}\underline{u}} \sim \frac{|E_2|}{|E_1+E_2|}$$

where E_1 and E_2 have independent distributions and

(4.5.8) $E_2 \sim \mathcal{W}_1(I,\nu+m-1:(0)) \sim \chi^2_{\nu-m+1}(0)$

(4.5.9($E_1 \sim \mathcal{W}_1(I,m:k) \sim \chi^2_m(\tfrac{1}{2}\underline{\mu}'V^{-1}\underline{\mu})$

since k is the "diagonal matrix" (here 1x1 and therefore scalar) such that k is equal to the non-zero latent root of $V^{-1}\underline{\mu}\,\underline{\mu}'$. Finally simplifying (4.5.7)

$$(4.5.10)\quad \underline{u}'W^{-1}\underline{u} \sim \frac{\chi^2_m(\tfrac{1}{2}\underline{\mu}'V^{-1}\underline{\mu})}{\chi^2_{\nu-m+1}(0)}$$

$$\sim \frac{m}{\nu-m+1}F_{m:\nu-m+1}(\tfrac{1}{2}\underline{\mu}'V^{-1}\underline{\mu}).$$

In particular if $\underline{\mu} = 0$, then the F-distribution becomes central.

This result is important in tests on the population mean vector (to be developed later) and is stated as Theorem 35 below.

Theorem 35

Let $\underline{u}$ and W have independent distributions such that

$$\underline{u} \sim N_m(\underline{\mu};V)$$

$$W \sim \mathcal{W}_m(V;\nu:(0)) \qquad \nu>m-1$$

then with $\lambda=\tfrac{1}{2}\underline{\mu}'V^{-1}\underline{\mu}$

$$p(\underline{u}'W^{-1}\underline{u}=x) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!}\,\frac{x^{\frac{1}{2}(2i+m)-1}(1+x)^{-\frac{1}{2}(2i+\nu+1)}}{B(\tfrac{1}{2}(\nu-m+1);\tfrac{1}{2}(2i+m))}$$

so that $\underline{u}'W^{-1}\underline{u} \sim \frac{m}{\nu-m+1}F_{m:\nu-m+1}(\lambda=\tfrac{1}{2}\underline{\mu}'V^{-1}\underline{\mu})$

4.6 Approximations to the distribution of the central U-statistic

The likelihood ratio test of the multivariate linear hypothesis is a function of a U-statistic. This U-statistic is an observation from a population of centrally distributed U-variates under the null hypothesis. The distribution of U in the multivariate test of linear hypotheses is as important therefore as its counterpart, the F-ratio in the univariate case.

Although we cannot write down a density for the scalar quantity U (except, as we shall see, in a few special cases) nevertheless, we have all moments of U expressed by equation (4.2.9) and using these moments, very adequate approximations to the cumulative density on U can be formulated. First, however, the exact distributions will be given.

It was seen (equation 4.2.13) that $U_{1:\nu_1:\nu_2}$ is related to the F variate. Using the relationship expressed in that equation, it is seen that

$$(4.6.1) \quad \Pr\{U_{1:\nu_1:\nu_2}<\tau\} = \Pr\left\{F_{\nu_1:\nu_2} > \frac{\nu_2}{\nu_1}\,\frac{1-\tau}{\tau}\right\}$$

thus percentage points for $U_{1:\nu_1:\nu_2}$ are available from those existing for the F-ratio. Alternatively by equation (4.2.15), it is evident that $U_{1:\nu_1:\nu_2}$ has a B-distribution so that

$$(4.6.2) \quad \Pr\{U_{1:\nu_1:\nu_2}<\tau\} = \int_0^{\tau} \frac{z^{\frac{1}{2}\nu_2-1}(1-z)^{\frac{1}{2}\nu_1-1}}{B(\frac{1}{2}\nu_2:\frac{1}{2}\nu_1)}\,dz$$

this integral (the Incomplete B-integral) is well tabulated .

The other special cases are obtained by comparing the moments of $U_{m:\nu_1:\nu_2}$ with the moments of products of certain random variables after the manner of equation (4.2.15).

Theorem 36

$$U_{2r:\nu_1:\nu_2} \sim \prod_{j=1}^{r} Y_j^2 \quad \text{where}$$

$\{Y_j\}_1^r$ are mutually independent and

$$p(Y_j) = \frac{Y_j^{\nu_2-2j}(1-Y_j)^{\nu_1-1}}{B(\nu_2-2j+1,\nu_1)}$$

Theorem 37

$$U_{2r+1:\nu_1:\nu_2} \sim \prod_{j=1}^{r} Y_j^2 X \quad \text{where}$$

$\{X:\{Y_j\}_1^r$ are mutually independent, Y_j is distributed as in Theorem 36 and

$$p(X) = \frac{X^{\frac{1}{2}(\nu_2+1-m)-1}(1-X)^{\frac{1}{2}\nu_1-1}}{B(\frac{1}{2}(\nu_2+1-m);\frac{1}{2}\nu_1)}$$

The proof of both Theorems rests in identifying and recognizing as equal the moments of the U-statistic and product of the functions of the B-variates.

By Theorem 36 we have immediately that

$$\text{(4.6.3)} \qquad \sqrt{U_{2:\nu_1:\nu_2}} \sim Y_1$$

and hence

$$(4.6.4) \qquad \Pr\{U_{2:\nu_1:\nu_2} < \tau\} = \int_0^{\sqrt{\tau}} \frac{z^{\nu_2-2}(1-z)^{\nu_1-1}dz}{B(\nu_2-1;\nu_1)} .$$

We may also write as a consequence of equation (4.6.4)

$$(4.6.5) \qquad U_{2:\nu_1:\nu_2} \sim U_{1:2\nu_1:2(\nu_2-1)}$$

so that

$$(4.6.6) \qquad \frac{1-\sqrt{U_{2:\nu_1:\nu_2}}}{\sqrt{U_{2:\nu_1:\nu_2}}} \times \frac{\nu_2-1}{\nu_1} \sim F_{2\nu_1:2\nu_2-2}$$

Percentage points for $U_{2:\nu_1:\nu_1}$ are evidently readily available.

Some other special cases are given in T. W. Anderson's Introduction to Multivariate Statistical Analysis; pp. 196-202. The cumulative density for $U_{3:\nu_1:\nu_2}$ can be expressed as the weighed sum of ν_2 untabulated integrals; the cumulative density for $U_{4:\nu_1:\nu_2}$ is expressible as the weighed sum of ν_2^2 such integrals. These expressions will not be reproduced here but their existance should be remembered when ν_2 is small. (In practice, ν_2 is often quite large, and the computations involved in obtaining the exact cumulative density is not worth the effort over comparitively simple approximations).

Rather than consider the distribution of $U_{m:\nu_1:\nu_2}$, we shall consider the distribution of

$$(4.6.7) \qquad \eta = -\rho \log U_{m:\nu_1:\nu_2}. \qquad (0 \leq \eta < \infty).$$

The moment generating function of η is

$$(4.6.8) \qquad \phi\eta(\theta) = e^{\theta\eta} = \{\log U_{m:\nu_1:\nu_2}\}^{-\theta\rho}$$

Using equation (4.2.9) the cumulant generating function for η is

$$(4.6.9) \qquad \Psi_\eta(\theta) = \log\phi_\eta(\theta)$$

$$= k + \sum_{j=1}^{m} \log \frac{\Gamma(\frac{1}{2}(\nu_2+1-j)-\rho\theta)}{\Gamma(\frac{1}{2}(\nu_1+\nu_2+1-j)-\rho\theta)}$$

where k has a value independent of θ. Rewriting this as

$$(4.6.10) \qquad \Psi_\eta(\theta) = k + \sum_{j=1}^{m} \log \frac{\Gamma(\alpha_j+\frac{\rho}{2}(1-2\theta))}{\Gamma(\beta_j+\frac{\rho}{2}(1-2\theta))}$$

where

$$\alpha_j = \tfrac{1}{2}(\nu_2+1-\rho-j)$$

$$(4.6.11)$$

$$\beta_j = \tfrac{1}{2}(\nu_1+\nu_2+1-\rho-j)$$

so that*

$$(4.6.12)\quad \Psi_\eta(\theta) = -\frac{m\nu_1}{2}\log(1-2\theta) + \sum_{r=1}^{\infty}\pi_r'\{\frac{1}{(1-2\theta)^r} - 1\}$$

with

$$(4.6.12)\quad \pi_r' = \frac{(-1)^r \sum_{j=1}^{m} B_{r+1}(\beta_j) - B_{r+1}(\alpha_j)}{r(r+1)(\rho/2)^r}$$

the constant being determined by $\Psi_\eta(\theta=0) = 0$.

Now $\sum_{j=1}^{m}\left\{B_s(\beta_j) - B_s(\alpha_j)\right\}$ is the coefficient of $\tau^s/s!$ in $G(\tau:\rho)$

where

$$(4.6.13)\quad G(\tau:\rho) =$$

---

*We use the expansion

$$\log\Gamma(x+h) = \log\sqrt{2\pi} + (x+h-\tfrac{1}{2})\log x$$

$$- x - \sum_{r=1}^{\infty}(-1)^r \frac{B_{r+1}(h)}{r(r+1)x^r}$$

where $B_s(h)$ is the s-th Bernoulli polynomial defined by

$$\tau e^{h\tau}\quad(e^{\tau}-1) = \sum_{s=0}^{\infty}\frac{\tau^s}{s!}B_s(h)$$

for example

$$B_1(h) = h-\tfrac{1}{2};\ B_2(h) = h^2-h+\frac{1}{6}$$

$$B_3(h) = h^3-\frac{3h^2}{2}+\tfrac{1}{2}h$$

$$\frac{\tau(e^{\frac{1}{2}\tau m}-1)}{(e^{\tau}-1)(e^{\frac{1}{2}\tau}-1)}\left(e^{\frac{1}{2}\tau(\nu_1+\nu_2+1-\rho-m)}-e^{\frac{1}{2}\tau(\nu_2+1-\rho-m)}\right)$$

and $G(\tau:\rho)$ is an odd function of τ when

$$(4.6.14)\qquad \rho = \rho_0 = \nu_2 - \tfrac{1}{2}(m+1-\nu_1).$$

Hence

$$(4.6.15)\qquad \phi_\eta(\theta)\Big|_{\rho=\rho_0} = \frac{1}{(1-2\theta)^{\frac{m\nu_1}{2}}}\exp\sum_{s=1}^{\infty}\pi_{2s}\left((1-2\theta)^{-2s}-1\right)$$

where

$$(4.6.16)\qquad \begin{aligned}\pi_2 &= \pi_2'\Big|_{\rho=\rho_0} = \frac{m\nu_1}{48\rho_0^2}(m^2+\nu_1^2-5)\\ \pi_4 &= \pi_4'\Big|_{\rho=\rho_0} = \frac{m\nu_1}{1920\rho_0^4}(3m^4+3\nu_1^4+10m^2\nu_1^2-50m^2-50\nu_1^2+159)\end{aligned}$$

Expanding the exponential in equation (4.6.15)

$$(4.6.17)\qquad \phi_\eta(\theta)\Big|_{\rho=\rho_0} = \frac{1-\pi_2-\pi_4-\frac{1}{2}\pi_2^2}{(1-2\theta)^{\frac{1}{2}m\nu_1}}$$

$$+\frac{\pi_2-\pi_2}{(1-2\theta)^{\frac{1}{2}m\nu_1+2}}$$

$$+\frac{\pi_4+\frac{1}{2}\pi_2^2}{(1-2\theta)^{\frac{1}{2}m\nu_1+4}} + R(\rho_0^{-6})$$

with $R(\rho_0^{-6})$ a remainder involving terms in $\rho_0^{-6}, \rho_0^{-8}, \ldots, \ldots$.
Finally therefore

$$(4.6.18) \quad Pr(-\rho_0 \log U_{m:\nu_1:\nu_2} \le U) =$$

$$(1-\pi_2-\pi_4-\pi_2^2)\, Pr(\chi^2_{m\nu_1} \le u)$$

$$+ \quad (\pi_2-\pi_2^2)\, Pr(\chi^2_{m\nu_1+4} \le u)$$

$$+ \quad (\pi_4+\pi_2^2)\, Pr(\chi^2_{m\nu_1+8} \le u) + R'(\rho_0^{-6})$$

with $R'(\rho_0^{-6})$ a remainder involving terms in $\rho_0^{-6}, \rho_0^{-8}, \ldots, \ldots$

Usually ρ_0 is of the order of the sample size in an analysis of the multivariate linear hypothesis and for these purposes the approximation

$$(4.6.19) \quad Pr(-\rho_0 \log U_{m:\nu_1:\nu_2} \le u) \doteqdot Pr(\chi^2_{m\nu_1} \le u)$$

is entirely adequate using the value of ρ_0 given in equation (4.6.14).

The central variate $U_{m:\nu_1:\nu_2}$ is contained in a class of random variables considered by Box (1949) and it is his approach which has been used to arrive at equation (4.6.18). Many of the distributional approximations in the multivariate test of hypothesis are a result of Box's work. Henceforth the results will be quoted with a reference to Box.

It is noticed that since $U_{m:\nu_1:\nu_2} \sim U_{\nu_1:m:\nu_1+\nu_2-m}$, equation (4.6.18) must be unaffected if

$$\begin{cases} m \text{ is replaced by } \nu_1 \\ \nu_1 \text{ is replaced by } m \text{ and} \\ \nu_2 \text{ is replaced by } \nu_1+\nu_2-m \text{ ;} \end{cases}$$

which provides a useful check on the result.

4.7 An inequality on the cumulative integral for the noncentral U-variate

It was shown earlier (Section 4.3) that if W_1 and W_2 have independent Wishart distributions with

$$\begin{aligned} W_1 &\sim \mathcal{W}_m(V,\nu_1\Lambda) \\ W_2 &\sim \mathcal{W}_m(V,\nu_2(0)) \end{aligned} \tag{4.7.1}$$

then

$$\frac{|W_2|}{|W_1+W_2|} \sim \frac{|C_2|}{|C_1+C_2|} \tag{4.7.2}$$

where

$$\begin{aligned} C_1 &\sim \mathcal{W}_m(I,\nu_1\Phi) \\ C_2 &\sim \mathcal{W}_m(I,\nu_2(0)) \end{aligned} \tag{4.7.3}$$

such that Φ is a diagonal matrix whose diagonal elements account for all the latent roots of Λ in the metric of V; that is $\{\phi_{ii}\}_{i=1}^{m}$ the solutions of $|\Lambda-\phi V| = 0$.

We now investigate the effect of changing ϕ_{ii} on the probability $\Pr\{|W_2|>\tau|W_1+W_2|\}$ and prove:-

Theorem 38

If W_1 and W_2 have independent distributions such that

$$W_1 \sim \mathcal{W}_m(V;\nu_1\Lambda)$$

$$W_2 \sim \mathcal{W}_m(V;\nu_2(0))$$

and if $\{\phi_{ii}\}_{i=1}^{m}$ are the latent roots of Λ in the metric of V, then $\Pr\{|W_2|>\tau|W_1+W_2|\}$, $0<\tau<1$, is a decreasing function of the "free"(*) ϕ_{ii}.

*It is to be remembered that Λ is a positive semi-definite (mxm) matrix of rank $\min(m,\nu_1)$ so that when $\nu_1<m$ at least $(m-\nu_1)$ of the set $\{\phi_{ii}\}_{i=1}^{m}$ must be zero. The remaining ν_1 ϕ_{ii} are "free" to take any value in $(0,\infty)$.

Without loss of generality we shall consider $\Pr\{|C_2|>\tau|C_2+C_1|\}$ and allow

$$C_1 = \sum_{j=1}^{\nu_1}\underline{u}_j\underline{u}_j' \tag{4.7.4}$$

with $\{\underline{u}_j\}_{j=1}^{\nu_1}$ a set of independent normal vectors such that

$$\underline{u}_j \sim N_m(\phi_{jj}^{\frac{1}{2}}\underline{\varepsilon}_j;I) \tag{4.7.5}$$

where

$$\underline{\varepsilon}_j' = (\delta_{1j},\delta_{2j},\ldots,\delta_{mj}),\quad \delta_{ij} = \begin{matrix} 0 & i\neq j \\ 1 & i=j \end{matrix} \tag{4.7.6}$$

thus $C_1 \sim \mathcal{W}_m(I;\nu_1 : \Phi)$ as required. We will allow ϕ_{11} to be a free latent root and write

$$C_1 = H + \underline{u}_1\underline{u}_1' \tag{4.7.7}$$

and note that H is positive semi-definite. It is convenient to introduce

Lemma 1

For fixed C_2 and H, the inequality $|C_2| > \tau |C_2 + H + \underline{u}_1'\underline{u}_1|$, $0<\tau<1$, represents the interior of an m-dimensional ellipse in the space of $\underline{u}_1$.

Proof:- Since $|C_2+H+\underline{u}_1\underline{u}_1'| = (1+\underline{u}_1'(C_2+H)^{-1}\underline{u}_1)|C_2+H|$ the inequality may be written

$$\underline{u}_1'(C_2+H)^{-1}\underline{u}_1 < \frac{|C_2|}{\tau|C_2+H|} - 1$$

which, since (C_2+H) is positive definite, is the interior of an m-dimensional ellipse, centre the origin, in the space of $\underline{u}_1$.

Lemma 2

Let E be a convex set, symmetric with respect to reflection through the origin and define $E_k(\underline{\zeta}) = \{\underline{x}+k\underline{\zeta} \mid \underline{x} \in E\}$. If K is a convex set symmetric with respect to the origin then for fixed $\underline{\zeta}$ volume $(K \cap E_k(\underline{\zeta}))$ is a non-increasing function of k.

Proof:-* The volume of $E_k(\underline{\zeta})$ is independent of k so that $\text{vol}\{E_1(\zeta)\} = \text{vol}\{E_{-1}(\underline{\zeta})\}$. Suppose $0<k<1$, then

$$E_k(\underline{\zeta}) = \frac{1+k}{2} E_1(\underline{\zeta}) + \frac{1-k}{2} E_{-1}(\underline{\zeta})$$

---

*See T. W. Anderson (1955: Proc. Amer. Math. Soc.)

that is, if $\underline{x}_{(1)}+\underline{\zeta}$ is a point of $E_1(\underline{\zeta})$ and $\underline{x}_{(-1)}-\underline{\zeta}$ is a point of $E_{-1}(\underline{\zeta})$, then

$$\frac{1+k}{2}(\underline{x}_{(1)}+\underline{\zeta}) + \frac{1-k}{2}(\underline{x}_{(-1)}-\underline{\zeta}) = (\frac{1+k}{2}\underline{x}_{(1)}+\frac{1-k}{2}\underline{x}_{(-1)})+k\underline{\zeta}$$

is a point of $E_k(\underline{\zeta})$ since $\frac{1+k}{2}\underline{x}_{(1)}+\frac{1-k}{2}\underline{x}_{(-1)}$ is necessarily a point of E by the convex property of that set. Consequently

$$K \cap E_k(\underline{\zeta}) \supset \left[\frac{1+k}{2} K \cap E_1(\underline{\zeta}) + \frac{1-k}{2} K \cap E_{-1}(\underline{\zeta})\right].$$

For if P is any point of $\frac{1+k}{2} K \cap E_1(\underline{\zeta}) + \frac{1-k}{2} K \cap E_{-1}(\underline{\zeta})$ then there exist

$$y_{(a)} \in K \cap E_1(\underline{\zeta}) \quad \text{(so that } y_{(a)} \in K;\ y_{(a)} \in E_1(\underline{\zeta}))$$

$$y_{(b)} \in K \cap E_{-1}(\underline{\zeta}) \quad \text{(so that } y_{(b)} \in K;\ y_{(b)} \in E_{-1}(\underline{\zeta}))$$

such that P is $\frac{1+k}{2} y_{(a)}+\frac{1-k}{2} y_{(b)}$. But since $y_{(a)} \in E_1(\underline{\zeta})$ and $y_{(b)} \in E_{(-1)}(\underline{\zeta})$, then $P \in E_k(\underline{\zeta})$. Also since $y_{(a)} \in K$ and $y_{(b)} \in K$ and K is convex, P belongs to K for $0<\frac{1+k}{2}=1-\frac{1-k}{2}<1$. Hence $P \in K \cap E_k(\underline{\zeta})$.

The Brunn-Minkowski Theorem states that for A and B non-empty convex sets and $0<\theta<1$, vol $(\theta A+(1-\theta)B) \geq \theta\text{vol}(A)+(1-\theta)\text{vol}(B)$. Hence

$$\text{vol}(K \cap E_k(\underline{\zeta})) \geq \frac{1+k}{2}\text{vol}(K \cap E_1(\underline{\zeta}))+\frac{1-k}{2}\text{vol}(K \cap E_{-1}(\underline{\zeta}))$$

and since, by the symmetry of K and E

$$\text{vol}(K \cap E_1(\underline{\zeta})) = \text{vol}(K \cap E_{-1}(\underline{\zeta}))$$

we have

$$\text{vol}(K \cap E_k(\underline{\zeta})) \geq \text{vol}(K \cap E_1(\underline{\zeta}))$$

It follows as a simple corrollary that $\mathrm{vol}(K \wedge E_k(\underline{\zeta}))$ is a non-increasing function of k for fixed $\underline{\zeta}$.

Lemma 3.

Let $f(\underline{x}) = f(-\underline{x})$ be a density on vector $\underline{x}$ and define $K_u : \{\underline{x} \mid f(\underline{x}) > u\}$. If K_u is convex and $E_k(\underline{\zeta})$ is as defined in Lemma 2, then

$$\int_{E_k(\underline{\zeta})} f(\underline{x})d\underline{u} \text{ decreases with k for any fixed } \underline{\zeta}.$$

for if $V_k(u) = \mathrm{vol}(E_k(\underline{\zeta}) \wedge K_u)$ then

$$\int_{E_k(\underline{\zeta})} f(\underline{x})d\underline{x} = \int_{u=f(\underline{0})}^{u=0} u\,dV_k(u) \qquad \text{so that, integrating by parts}$$

$$\int_{E_k(\underline{\zeta})} f(\underline{x})d\underline{x} = \Big[uV_k(u)\Big]_{u=f(\underline{0})}^{u=0} + \int_{u=0}^{u=f(\underline{0})} V_k(u)du$$

The first term on the right-hand side is zero since $V_k(0)=\mathrm{vol}E_k(\underline{\zeta})$ is finite and $V_k(u=f(\underline{0}))=0$; the second term is a decreasing function of k.

Proof of Theorem 38

$$\Pr[|W|_2 > \tau|W_1+W_2|] = \mathscr{E}_{C_2}\,\mathscr{E}_{H} \int_E \frac{e^{-\frac{1}{2}(\underline{u}_1-\phi_{11}^{\frac{1}{2}}\underline{\varepsilon}_1)'(\underline{u}_1-\phi_{11}^{\frac{1}{2}}\underline{\varepsilon}_1)}}{(2\pi)^{\frac{1}{2}m}}\, d\underline{u}_1$$

where by Lemma 1, E is a hyper-ellipse (in the space of $\underline{u}_1$) centre the origin. Writing $\underline{x} = \underline{u}_1 - \phi_{11}^{\frac{1}{2}}\underline{\varepsilon}_1$,

$$\Pr[|W_2|>\tau|W_1+W_2|] = \underset{C_2}{\mathcal{E}}\,\underset{H}{\mathcal{E}} \int_{E^{\frac{1}{2}}_{\phi_{11}}(\underline{\varepsilon}_1)} \frac{e^{-\frac{1}{2}\underline{x}'\underline{x}}}{(2\pi)^{\frac{1}{2}m}}\, d\underline{x}$$

By Lemma 3, the integral is a decreasing function of $\phi^{\frac{1}{2}}_{11}$ for fixed C_2 and H. Finally taking expectations over C_2 and H we complete a proof of the theorem.

4.8 The "step-down" property of $U'_{m:\nu_1:\nu_2}$

Let

(4.8.1) $E \sim \mathcal{W}_m(V;\nu_2:(0))$

and

(4.8.2) $H = ZZ' \sim \mathcal{W}_m(V;\nu_1:\Lambda)$

where

(4.8.3) $Z \sim N_{m\times\nu_1}(\Gamma;V\otimes I)$

and (4.8.4) $\Gamma\Gamma' = \Lambda$.

Partition E as

(4.8.5) $E = \left[\begin{array}{c|c} E_{11} & E_{12} \\ \hline E_{21} & E_{22} \end{array}\right]$

with E_{11} a pxp matrix and E_{22} a qxq matrix; and let V, H and Λ be similarly partitioned. Partition Γ as

(4.8.6) $$\Gamma = \begin{bmatrix} \Gamma_{(1)} \\ \hdashline \Gamma_{(2)} \end{bmatrix} ;\ \left(\begin{smallmatrix} p \\ q \end{smallmatrix} \text{-} \times \nu_1\right)$$

It was shown (Section 3.6) that $|E| / |E_{11}|$ is distributed independently of $|E_{11}|$ (Theorem 19). Also from the theory proceeding Theorem 19, $|E+H|$ and $|E+H|/|E_{11}+H_{11}|$ have independent distributions.

Now

(4.8.7) $$\frac{|E|}{|E_{11}|} \sim \mathcal{W}_q(V_{22}-V_{21}V_{11}^{-1}V_{12};\nu_2-p:(0))$$

(4.8.8) $$\frac{|E+H|}{|E_{11}+H_{11}|} \sim \mathcal{W}_q(V_{22}-V_{21}V_{11}^{-1}V_{12};\nu_1+\nu_2-p:(0)) \quad \underline{\underline{\text{if}}}$$

$$\Gamma_{(2)} = V_{21}V_{11}^{-1}\Gamma_{(1)}$$

Although $|E| / |E_{11}|$ and $|E+H| / |E_{11}+H_{11}|$ are not independent, their ratio can be shown to have the distribution of $U_{q}:\nu_1:\nu_2-p$ (central) whenever $\Gamma_{(2)} = V_{21}V_{11}^{-1}\Gamma_{(1)}$ holds. This result leads to an important theorem.

Theorem 39

Let E and Z be independently distributed with $E \sim \mathcal{W}_m(V;\nu_2:(0))$ and $Z \sim N_{m\times\nu_1}(\Gamma;V \otimes I)$ and write $H=ZZ'$ so that $H \sim {}_m(V;\nu_1:\Gamma\Gamma')$. With E,H,V and Γ partitioned as in the text, then whenever $\Gamma_{(2)}=V_{21}V_{11}^{-1}\Gamma_{(1)}$,

$$\frac{|E_{11}|}{|E_{11}+H_{11}|} \quad \text{and} \quad \frac{|E|/|E_{11}|}{|E+H|/|E_{11}+H_{11}|} \quad \text{have independent}$$

distributions. Further

$$\frac{|E_{11}|}{|E_{11}+H_{11}|} \sim U'_{p:\nu_1:\nu_2} \qquad (\text{central iff } \mu_{(1)}=(0))$$

and

$$\frac{|E|/|E_{11}|}{|E+H|/|E_{11}+H_{11}|} \sim U_{q:\nu_1:\nu_2-p}.$$

In a proof of Theorem 39 we shall assume that ν_2 is an integer and allow E to be represented as $E = YY'$ where $Y \sim N_{m\times\nu_2}((0);V \otimes I)$. Now

$$(4.8.9) \quad \frac{|E|}{|E_{11}|} = \left|\left(Y_{(2)} \vdots Z_{(2)}\right)\left(\begin{array}{c|c} A & (0) \\ \hline (0) & (0) \end{array}\right)\left(\begin{array}{c} Y'_{(2)} \\ \hline Z'_{(2)} \end{array}\right)\right|$$

and

$$(4.8.10) \quad \frac{|E+H|}{|E_{11}+H_{11}|} = \left|\left(Y_{(2)} \vdots Z_{(2)}\right) B \left(\begin{array}{c} Y'_{2} \\ \hline Z'_{12} \end{array}\right)\right|$$

where

$$(4.8.11) \quad A = I - Y'_{(1)}\left(Y_{(1)}Y'_{(1)}\right)^{-1}Y_{(1)}$$

$$(4.8.12) \quad B = I - (Y_{(1)} \vdots Z_{(1)})'[(Y_{(1)} \vdots Z_{(1)})(Y_{(1)} \vdots Z_{(1)})']^{-1}(Y_{(1)} \vdots Z_{(1)}).$$

Clearly A and B are idempotent. Upon investigation it is found that

$$B - \left(\begin{array}{c|c} A & (0) \\ \hline (0) & (0) \end{array}\right) \text{ is idempotent.}$$

An application of Theorem 15 and the definition of the central U-variate completes the proof.

An immediate consequence of Theorem 39 is:

Theorem 40

Let $E \sim \mathcal{W}_m(V;\nu_2:(0))$ and $H \sim \mathcal{W}_m(V;\nu_i:(0))$; then if

$$\alpha_r = \begin{vmatrix} e_{11} & e_{12} \cdots & e_{1r} \\ e_{21} & e_{22} \cdots & e_{2r} \\ \vdots & \vdots & \vdots \\ e_{r1} & e_{r2} & e_{rr} \end{vmatrix} \quad \text{and} \quad \beta_r = \begin{vmatrix} e_{11}+h_{11} & e_{12}+h_{12} \cdots & e_{1r}+h_{1r} \\ e_{21}+h_{21} & e_{22}+h_{22} \cdots & e_{2r}+h_{2r} \\ \vdots & \vdots & \vdots \\ e_{r1}+h_{r1} & e_{r2}+h_{r2} \cdots & e_{rr}+h_{rr} \end{vmatrix}$$

then $\left\{ \frac{\alpha_1}{\beta_1} \, , \left\{ \frac{\alpha_r/\alpha_{r-1}}{\beta_r/\beta_{r-1}} \right\}_{r=2}^{m} \right\}$ are a set of independently distributed variates such that

$$\frac{\alpha_1}{\beta_1} \sim U_{1:\nu_1:\nu_2}$$

$$\frac{\alpha_r/\alpha_{r-1}}{\beta_r/\beta_{r-1}} \sim U_{1:\nu_1\nu_2 r+1} \qquad r = 2, 3, \ldots, m.$$

It is noted that this result is in keeping with equation (4.2.15) since

$$\text{(4.8.13)} \qquad \frac{|E|}{|E+H|} = \frac{\alpha_1}{\beta_1} \prod_{r=2}^{m} \frac{\alpha_r/\alpha_{r-1}}{\beta_r/\beta_{r-1}} \sim \prod_{r=1}^{m} U_{1:\nu_1:\nu_2-r+1}$$

the product in the last equation representing a product of m independently distributed U-variates with degrees of freedom as indicated.

CHAPTER 5

MOMENT GENERATING FUNCTIONS

5.1 Introduction

In this section we will obtain the joint moment generating function for

(a) a normally distributed vector and the normal density.

(b) a non-central Wishart variate with integer degrees of freedom.

(c) the central Wishart density with non-integer degrees of freedom.

The joint characteristic functions may be obtained but it turns out that, in the case of the densities listed above, the characteristic function is simply the moment generating function with complex argument. Since the derivation of the characteristic function requires some results from the theory of complex integration (which theory may require some revision on the part of the reader) we choose to omit such derivations.

5.2 The moment generating function of a normally distributed vector.

After the manner of section 2.7 we will define

(5.2.1) $\underline{y} = \underline{\mu} + G\underline{w}$

where

(5.2.2) $\underline{w} \frown N_m(\underline{0};I)$

and G is any kxm matrix of constants

so that

$$\underline{y} \sim N_k(\underline{\mu};\ GG'). \tag{5.2.3}$$

The moment generating function of the normally distributed vector is defined by

$$\Phi_{\underline{y}}(\underline{\theta}) = \mathcal{E}\, e^{\underline{\theta}'\underline{y}} \tag{5.2.4}$$

and it is noted that even though $\underline{y}$ does not have a density (unless $|GG'| \neq 0$) nevertheless functions of the elements of $\underline{y}$ have well-defined expectation. Using (5.2.1)

$$\Phi_{\underline{y}}(\underline{\theta}) = e^{\underline{\theta}'\underline{\mu}} \cdot \mathcal{E}\, e^{\underline{\theta}'G\underline{w}} \tag{5.2.5}$$

with density of $\underline{w}$ given by (5.2.2), hence

$$\Phi_{\underline{y}}(\underline{\theta}) = e^{\underline{\theta}'\underline{\mu}-\underline{\theta}'GG'\underline{\theta}} \int \cdots \int_{\{-\infty<w_i<+\infty\}_1^m} \frac{1}{(2\pi)^m}\, e^{-\frac{1}{2}(\underline{w}-G'\underline{\theta})'(\underline{w}-G'\underline{\theta})} \quad d\underline{w} \tag{5.2.6}$$

Recognizing the integral as the integral over a normal m-variate density, it follows that

$$\Phi_{\underline{y}}(\underline{\theta}) = \exp(\underline{\theta}'\underline{\mu}+\underline{\theta}'GG'\underline{\theta}) \qquad (\text{any } \underline{\theta}). \tag{5.2.7}$$

It is noted that if $|GG'| = 0$, there exist a non-trivial $\underline{a}$ such that $\underline{a}'G = \underline{0}'$. Writing $\underline{\theta} = \theta\underline{a}$, we obtain $\Phi_{\underline{a}'\underline{y}}(\theta) = \exp(\theta\underline{a}'\underline{\mu})$; that is $\underline{a}'\underline{y}$ takes the value $\underline{a}'\underline{\mu}$ with probability one.

The moment generating function for the normal m-variate density is

(5.2.8) $$\Phi_{\underline{x}}(\underline{\theta}) = \int\cdots\int \frac{|V|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}m}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'V^{-1}(\underline{x}-\underline{\mu})} e^{\underline{\theta}'\underline{x}} \quad d\underline{x}$$

and it is easy to show that

(5.2.9) $$\Phi_{\underline{x}}(\underline{\theta}) = \exp(\underline{\theta}'\underline{\mu}+\underline{\theta}'V\underline{\theta}).$$

5.3 The moment generating function of a noncentrally distributed Wishart variate.

We shall suppose

(5.3.1) $$X \frown \mathcal{N}_{mx}(M;\ V \otimes I)$$

and, after the manner of Section 3.2, define W by

(5.3.2) $$W = XX'.$$

With θ a symmetric mxm matrix, we define the moment generating function of the elements of W by

(5.3.3) $$\Phi_W(\theta) = \exp\sum_{i=1}^{m}\sum_{j=1}^{m} W_{ij}\theta_{ij}$$

$$= \exp \operatorname{tr}(\theta W).$$

Now we may write the density for $X = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_3)$ as

(5.3.4) $$f(x) = \frac{|V|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}m\nu}} e^{-\frac{1}{2}\operatorname{tr}(V^{-1}(X-M)(X-M)')}$$

and recognize that this represents a density for any positive definite V and any M. Now

$$\Phi_W(\Theta) = \int\cdots\int \frac{|V|^{-\frac{1}{2}\nu}}{(2\pi)^{\frac{1}{2}m\nu}} e^{-\frac{1}{2}\mathrm{tr}\, V^{-1}(XX'-X\mu'-\mu X'+\mu\mu'-2V\Theta XX')}\, dX \tag{5.3.5}$$

the integral being finite wherever $V^{-1}-2\Theta$ is positive definite. Writing $H = (V^{-1}-2\Theta)^{-1}$ for convenience we have

$$\Phi_W(\Theta) = \frac{|H|^{\frac{1}{2}\nu}}{|V|} \times e^{\frac{1}{2}\mathrm{tr}\, Q} \times \int\cdots\int \frac{|H|^{-\frac{1}{2}\nu}}{(2\pi)^{\frac{1}{2}m\nu}} e^{-\frac{1}{2}\mathrm{tr}\, H^{-1}(X-\Xi)'(X-\Xi)}\, dX \tag{5.3.6}$$

where

$$\begin{aligned} \Xi &= HV^{-1}\mu \\ Q &= \mu'\left\langle V^{-1}HV^{-1}-V^{-1}\right\rangle\mu. \end{aligned} \tag{5.3.7}$$

The integral in (5.3.5) is unity by comparison with $N_{m\times\nu}(\Xi;\, H \otimes I)$ whenever H is positive definite so that

$$\Phi_W(\Theta) = \frac{e^{-\frac{1}{2}\mathrm{tr}[(I-2\Theta V)^{-1}-I]V^{-1}\mu\mu'}}{|I-2\Theta V|^{\frac{1}{2}\nu}} \tag{5.3.8}$$

and the moment generating function is a function of V and $\mu\mu'$.

The moment generating function for the central Wishart variate is

$$\Phi_{\text{central } W}(\Theta) = |I-2\Theta V|^{-\frac{1}{2}\nu} \tag{5.3.9}$$

It is perhaps worth emphasizing that equation (5.3.6) gives the moment generating function of the non-central Wishart variate for v a positive integer since the approach has been to represent W as ZZ' with Z an $m \times v$ matrix having the density $N_{m\times v}((0); V \times I)$. If v is non-integer we cannot, of course, make use of this representation of W.

5.4 The moment generating function of the Wishart distribution with possibly non-integer degrees of freedom.

Again we obtain $\exp \sum_i \sum_j \Theta_{ij} W_{ij}$ with $\Theta = (\Theta_{ij})$ an $m\times m$ symmetric matrix when we shall suppose

$$\left.\begin{aligned} f(W) &= \frac{|V|^{-\frac{1}{2}v}\,|W|^{\frac{1}{2}(v-m-1)}}{2^{\frac{1}{2}mv}\quad \pi^{\frac{1}{4}m(m-1)}}\;\frac{e^{-\frac{1}{2}\mathrm{tr}(V^{-1}W)}}{\prod_{j=1}^{m}\Gamma(\frac{v+1-j}{2})}\,, && W \text{ pos. def.}\\ &= 0 &&, \quad \text{otherwise} \end{aligned}\right\}$$

with $v>m-1$.

$$(5.4.1) \qquad \phi_W(\Theta) = \underset{\text{pos. def. } W}{\int\!\!\ldots\!\!\int} \frac{|V|^{-\frac{1}{2}v}\quad |W|^{\frac{1}{2}(v-m-1)}}{2^{\frac{1}{2}mv}\qquad \pi^{\frac{1}{4}m(m-1)}}\;\frac{e^{-\frac{1}{2}\mathrm{tr}\; W(V^{-1}-2\Theta)}}{\prod_{j=1}^{m}\Gamma(\frac{v+1-j}{2})}$$

Now if $(V^{-1}-2\Theta)$ is positive definite, we could regard $|V|^{\frac{1}{2}}\phi_W(\Theta)$ $|V^{-1}-2\Theta|^{\frac{1}{2}v}$ on the integral over positive definite W of the density $W_m(V^{-1}-2\Theta)^{-1};\ v:(0))$. Hence

$$(5.4.2) \qquad \phi_W(\Theta) = \frac{1}{|I-2\Theta V|^{\frac{1}{2}v}}$$

which is exactly the result of (5.3.7).

It is noted that the characteristic function can be used to prove Theorem 12 for non-integer degrees of freedom.

CHAPTER 6

LATENT ROOTS AND VECTORS ASSOCIATED WITH WISHART VARIATES

6.1 Introduction

There are a series of tests; alternative to maximum likelihood tests; of the general linear hypothesis; which are based on a consideration of the latent roots of an observed Wishart variate in the metric of another. These alternative tests are not especially attractive either computationally or in respect of their distributional properties. It cannot be said of them, however, that they are less attractive than the likelihood ratio tests so that the problem of latent roots has received considerable attention in the modern theory of multivariate analysis. We will present in this chapter some of the main results connected with the distribution of latent roots. It is not our intention to pursue the matter too deeply, however, since it is the purpose of these notes to develop the theory and techniques connected with likelihood ratio tests of hypothesis.

It was noted in the introduction to Chapter 4 that the likelihood test of the general linear hypothesis is based on a statistic of the form $|E|/|E+H|$. If the null hypothesis be true, both E and H have Wishart distributions and are independent. Now

$$\frac{|E|}{|E+H|} = \frac{1}{|I+E^{-1}H|} = \prod_{j=1}^{m} \frac{1}{(1+\theta_j)}$$

where $(\theta_j)_1^m$ are the latent roots of $E^{-1}H$; that is $(\theta_j)_1^m$ are the solutions of $|H-\theta E| = 0$ and are thus described as the latent roots of H in the metric of E. It is noted that if $H \sim W_m(V;\nu_1:\Lambda)$ then $\max(0;m-\nu_1)$ of the latent roots are zero. It is further noted that if $H = \underline{u}\underline{u}'$ with $\underline{u}$ $N_m(\mu,V)$ and independent of E, then

$$\frac{|E|}{|E+\underline{u}\underline{u}'|} = \frac{1}{1+\underline{u}'E^{-1}\underline{u}} = \frac{1}{1+\theta}$$

with θ the sole non-zero root of $|\underline{u}\underline{u}'-\theta E| = 0$. So that $\theta = \underline{u}'E^{-1}\underline{u}$ has a distribution of the Hotelling's T^2 type.

The tests of the latent root type consist in assessing the departure of an observed set $(\theta_j)_1^m$ from expected value. Typical of such tests are

$$\max_j[(\theta_j)_1^m] \qquad \text{(Roy, S. V. 1953)}$$

6.2 The distribution of latent roots.

In the proofs and discussion in this section we shall assume

$$(6.2.1) \qquad \begin{cases} E \sim \mathcal{W}_m(I;\nu_2:(0)) \\ H \sim \mathcal{W}_m(I;\nu_1:(0)) \end{cases}$$

holding therefore that both E and H are non-singular. It is clear that there is no loss of generality in assuming the identity matrix for the dispersion matrix of both Wishart variates (compare section 4.3); the case $\nu_1<m$ is discussed later.

We consider solutions of

$$(6.2.2) \qquad |H-\phi(E+H)| = 0$$

so that

(6.2.3) $$\phi_i = \frac{\theta_i}{1+\theta_i} \quad ; \quad \theta_i = \frac{\phi_i}{1-\phi_i} \; .$$

It is convenient to label the $(\phi_i)_1^m$ in order of magnitude; thus, $1>\phi_1>\phi_2>\ldots>\phi_{m-1}>\phi_m$ (>0) and write

(6.2.4) $$D_\phi = \begin{bmatrix} \phi_1 & 0 & 0 \; . \; . \; . \; 0 \\ 0 & \phi_2 & 0 \; . \; . \; . \; 0 \\ 0 & 0 & \phi_3 \; . \; . \; . \; 0 \\ \cdot & \cdot & \cdot \qquad\quad \cdot \\ \cdot & \cdot & \cdot \qquad\quad \cdot \\ 0 & 0 & 0 \; . \; . \; . \; \phi_m \end{bmatrix}$$

If $\underline{g}_j$ be the vector (unique, apart from scale, with probability one) associated with ϕ_j then

(6.2.5) $$\left(H-\phi_j(E+H)\right)\underline{g}_j = \underline{0}$$

and we choose the scale, apart from the sign, by requiring

(6.2.6) $$\underline{g}_j'(E+H)\underline{g}_j = 1$$

Now

(6.2.7) $$\underline{g}_i'H\underline{g}_j = \left.\begin{array}{l} \phi_j\underline{g}_i'(E+H)\underline{g}_j \\ \\ \phi_i\underline{g}_i'(E+H)\underline{g}_j \end{array}\right\} \qquad i\neq j$$

hence, since $\phi_i = \phi_j$ $(i\neq j)$ with probability zero:

(6.2.8) $$\underline{g}_i'(E+H)\underline{g}_j = 0$$

Write

(6.2.9) $$G = (\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_m)$$

then

(6.2.10) $$G'(\Sigma+H)G = I$$

With

(6.2.11) $$K = G^{-1}$$

then

(6.2.12) $$\begin{aligned} H &= K'D_\phi K \\ \Sigma &= K'(I-D_\phi)K \end{aligned}$$

and we obtain the joint density of D and K.

The Jacobian $\left|\left|\frac{\partial(H;\Sigma)}{\partial(K;D_\phi)}\right|\right|$ is troublesome to obtain and rather than complicate this section, we give the derivation in Appendix B, simply quoting the result here:

(6.2.13) $$\frac{\partial(H;\Sigma)}{\partial(K;D_\phi)} = 2^m \|K\|^{m+2} \prod_{1<i<j<m}\prod (\phi_i-\phi_j).$$

Using (6.2.12), (6.2.13) and the joint densities of independent H and Σ,

(6.2.14) $$f(K,D_\phi) = \frac{2^m|K'K|^{\frac{1}{2}(m+2)}|K'D_\phi K|^{\frac{1}{2}(v_1-m-1)}}{2^{\frac{1}{2}m(v_1+v_2)}\ \pi^{\frac{1}{2}m(m-1)}} \times$$

$$\frac{|K'(I-D_\phi)K|^{\frac{1}{2}(v_2-m-1)}e^{-\frac{1}{2}\mathrm{tr}K'K}}{\prod_{j=1}^{m}\Gamma\left(\frac{v_1+1-j}{2}\right)\Gamma\left(\frac{v_2+1-j}{2}\right)} \prod_{i<j}\prod(\phi_i-\phi_j)$$

Simplifying:

$$(6.2.15) \qquad f(K,D_\phi) = \frac{2^m |K'K|^{\frac{1}{2}(\nu_1+\nu_2-m)} e^{-\frac{1}{2}\mathrm{tr}(K'K)}}{2^{\frac{1}{2}m(\nu_1+\nu_2)} \pi^{\frac{1}{2}m(m-1)} \prod_{j=1}^{m} \Gamma\left(\frac{\nu_1+1-j}{2}\right)\Gamma\left(\frac{\nu_2+1-j}{2}\right)} \times$$

$$\prod_i \phi_i^{\frac{1}{2}(\nu_1-m-1)} (1-\phi_i)^{\frac{1}{2}(\nu_2-m-1)} \prod_{i<j}\prod (\phi_i-\phi_j)$$

with $0<\phi_m<\phi_{m-1}<\ldots<\phi_2<\phi_1$

It is noted that the sign of each vector $\underline{g}_j$ was not fixed; this results in an ambiguity on K. Let us choose the K which has positive elements for its first row so that

$$(6.2.16) \qquad \begin{cases} 0<k_{1j}<\infty & j = 1, 2, \ldots, m \\ -\infty<k_{ij}<\infty & i = 2, 3, \ldots, m \\ & j = 1, 2, \ldots, m \end{cases}$$

Now

$$(6.2.17) \qquad 2^m \underset{\substack{(0<k_{1j}<\infty)_{j=1}^{m} \\ (-\infty<k_{ij}<\infty)_{i=2;\, j=1}^{m;\ m}}}{\int\cdots\int} |K'K|^{\frac{1}{2}(\nu_1+\nu_2-m)} e^{-\frac{1}{2}\mathrm{tr}(K'K)}$$

$$= \underset{(-\infty<k_{ij}<\infty)_{1;1}^{m;m}}{\int\cdots\int} |K'K|^{\frac{1}{2}(\nu_1+\nu_2-m)} e^{-\frac{1}{2}\mathrm{tr}(K'K)}$$

$$. = (2\pi)^{\frac{1}{2}m^2} \underset{K}{\mathcal{E}} |K'K|^{\frac{1}{2}(\nu_1+\nu_2-m)} \; ; \left(K \sim N_{m\times m}((0);I\times I)\right)$$

$$= (2\pi)^{\frac{1}{2}m^2} \underset{C}{\mathcal{E}} |C|^{\frac{1}{2}(\nu_1+\nu_2-m)} \; ; \left(C \sim \mathcal{W}_m(I;m:(0))\right)$$

$$= (2\pi)^{\frac{1}{2}m(\nu_1+\nu_2)} \prod_{j=1}^{m} \frac{\Gamma\left(\frac{\nu_1+\nu_2+1-j}{2}\right)}{\Gamma\left(\frac{m+1-j}{2}\right)}$$

Square matrix K and D_ϕ have independent distributions. Transforming from $(\phi_j)_1^m$ to $(\theta_j)_1^m$ we have the following theorems.

Theorem 41

If $W_{(i)} \sim \mathcal{W}_m(V;\nu_i:(0))$, $i = 1, 2$, and $W_{(1)}$ and $W_{(2)}$ are independent, then the solutions $(\theta_i)_1^m$ of

$$|W_{(1)}-\theta W_{(2)}| = 0$$

when $\nu_i \geq m$

have the joint density

$$f(\theta_1,\theta_2,\ldots,\theta_m) = \pi^{\frac{1}{2}m} \prod_{j=1}^{m} \frac{\Gamma\left(\frac{\nu_1+\nu_2+1-j}{2}\right)}{\Gamma\left(\frac{\nu_1+1-j}{2}\right)\Gamma\left(\frac{\nu_2+1-j}{2}\right)\Gamma\left(\frac{m+1-j}{2}\right)} \times$$

$$\prod_{j=1}^{m} \frac{\theta_j^{\frac{1}{2}(\nu_1-m-1)}}{(1+\theta_j)^{\frac{1}{2}(\nu_1+\nu_2)}} \prod_{1<i<j<m} (\theta_i-\theta_j)$$

$$\infty > \theta_1 > \theta_2 > \ldots > \theta_m > 0$$

Theorem 42

If $W_{(1)}$, $W_{(2)}$, and $\{\theta_j\}_1^m$ are as described in the previous theorem and if $\{\underline{g}_j\}_1^m$ are such that

$$(W_{(1)}-\theta_j W_{(2)})\underline{g}_j = 0$$

then $\{\underline{g}_j\}_1^m$ is distributed independently of $\{\theta_i\}_1^m$.

Further, if K is defined by

$$(\underline{g}_1, \underline{g}_2, \dots, \underline{g}_m)'(W_{(1)}+W_{(2)})(\underline{g}_1, \underline{g}_2, \dots, \underline{g}_m) = I$$

$$K = (\underline{g}_1, \underline{g}_2, \dots, \underline{g}_m)^{-1}$$

$$k_{1j} \geq 0 \qquad j = 1, 2, \dots, m.$$

then

$$f(K) = \prod_{j=1}^{m} \frac{\Gamma\left(\frac{m+1-j}{2}\right)}{\Gamma\left(\frac{v_1+v_2+1-j}{2}\right)} \cdot \frac{|K|^{v_1+v_2-m}}{2^{\frac{1}{2}m(v_1+v_2-2)}\pi^{\frac{1}{2}m^2}} \quad e^{-\frac{1}{2}\mathrm{tr}(K'K)}$$

$$0<k_{1j}<\infty \qquad j = 1, 2, \dots, m$$

$$-\infty<k_{ij}<\infty \qquad 2\ i;\ j = 1, 2, \dots, m.$$

The density of K given in Theorem 42 is closed related to the Haar Invariant density [Halmos 1950]. It can be shown that if

$$f(L) = \prod_{j=1}^{m} \frac{\Gamma(\frac{1}{2}(m+1-j))}{\Gamma(\frac{1}{2}(v_1+v_2+1-j)} \frac{|L'L|^{\frac{1}{2}(v_1+v_2-m)}}{2^{\frac{1}{2}m(v_1+v_2)}\ \pi^{\frac{1}{2}m^2}} \quad e^{-\frac{1}{2}\mathrm{tr}(L'L)}$$

$$\{-\infty<\ell_{ij}<\infty\}_{i=1;\,j=1}^{m;\ m}$$

then

$$f(L = K \mid \{0 < \ell_{1j} < \infty\}_{j=1}^{m}) \text{ is the density of Theorem 42.}$$

The Haar Invariant density has some interesting properties, the most obvious of which is that for any orthagonal P, $f(PL=L^*) = f(L=L^*)$. It may also be noted that

$$f(L \mid \nu_1+\nu_2 = m) \text{ is as } N_{m\times m}((0);\ I \times I).$$

Other simple properties are inspected in the exercises to this chapter.

We turn now to the case where H is a singular Wishart matrix and use the theory of section 4, wherein it is shown that

$$U_{m:\nu_1:\nu_2} \sim U_{\nu_1:m:\nu_1+\nu_2-m}. \qquad (6.2.18)$$

We use this result to show

Theorem 43

If $W_{(2)}$ ${}_{m}(V;\nu_2:(0))$, $\nu_2 > m-1$ and $W_{(1)}$ ${}_{m}(V;\nu_1:(0))$, $\nu_1 < m$ and integer then the ν_1 non-zero solutions of

$$|W_{(1)} - \theta W_{(2)}| = 0$$

have the joint density:

$$f(\theta_1,\theta_2,\ldots,\theta_{\nu_1}) = \pi^{\frac{1}{2}\nu_1} \prod_{j=1} \frac{\Gamma\left(\frac{\nu_1+\nu_2+1-j}{2}\right)}{\Gamma\left(\frac{\nu_1+1-j}{2}\right)\Gamma\left(\frac{\nu_1+\nu_2-m+1-j}{2}\right)\Gamma\left(\frac{m+1-j}{2}\right)} \times$$

$$\prod_{j=1}^{\nu_1} \frac{\theta_j^{\frac{1}{2}(m-\nu_1-1)}}{(1+\theta_j)^{\frac{1}{2}(\nu_1+\nu_2)}} \prod\prod_{1<i<j<\nu_1} (\theta_i-\theta_j)$$

$$\infty>\theta_1>\theta_2>\ldots>\theta_{\nu_1}>0.$$

It is noted that if we replace

$$\begin{cases} m \text{ by } \nu_1 \\ \nu_1 \text{ by } m \\ \nu_2 \text{ by } \nu_1+\nu_2-m \end{cases}$$

in the density for $(\theta_j)_1^m$ given in Theorem 41, we arrive at the density for $(\theta_j)_1^{\nu_1}$ given in Theorem 43.

A final result in this series can also be obtained as a consequence of Theorem 41.

Theorem 44

If $W \sim \mathcal{W}_m(V;\nu:(0))$ $(\nu>m-1)$ and $(\theta_j)_1^m$ are the solutions of

$$|W-\theta V| = 0$$

then

$$f(\theta_1,\theta_2,\ldots,\theta_m) = \frac{\pi^{\frac{1}{2}m}}{2^{\frac{1}{2}m}} \prod_{j=1}^{m} \frac{\theta_j^{\frac{1}{2}(\nu-m-1)}\, e^{-\frac{1}{2}\theta_j}}{\Gamma\left(\frac{m+1-j}{2}\right)\, \Gamma\left(\frac{\nu+1-j}{2}\right)} \prod_{1<i<j<m}\prod (\theta_i-\theta_j)$$

$$\infty>\theta_1>\theta\,,\ldots>\theta_m>0.$$

The density in Theorem 44 is obtained by considering

$$|W_{(1)}-(\nu_2\lambda)\frac{1}{\nu_2}W_{(2)}| = 0$$

and writing $\theta = v_2\lambda$. As v_2 approaches infinity, $\frac{1}{v_2} V_{(2)}$ approaches V so that

$$\left\{ f(\theta_1,\theta_2,\ldots,\theta_n) \right\}\Big\downarrow_{\text{Theorem 44}} = \left\{ \underset{v_2\to\infty}{\text{limit}}\ f(\lambda_1,\lambda_2,\ldots,\lambda_n)\Big|_{(\lambda_j=\frac{1}{v_2}\theta_j)} \right\}\Big\downarrow_{\text{Theorem 41}}$$

CHAPTER 7

THE GENERAL LINEAR HYPOTHESIS

7.1 Introduction

In the univariate case we assume that the expected value of a random variable x is a linear combination of a number of error-free variables and write

$$\mathcal{E}(x) = \beta_1\xi_1 + \beta_2\xi_2 + \ldots + \beta_\ell\xi_\ell$$

where $\{\xi_i\}_1^\ell$, are the known error-free variables and the $\{\beta_i\}_1^\ell$ are (usually unknown) multipliers. We may wish, in a regression experiment, to describe the average growth of a plant in terms of its age so that ξ_1 may be age in days, ξ_2 the square of the age in days and so on. In a design experiment we may set $\xi_j=1$ and $\xi_i=0$, all $i\neq j$ when an observation belongs to the j^{th} group or population. In both cases $\{\xi_i\}_1^\ell$ are error-free variables, that is variables which may be manipulated at the will of the experimenter.

We may write the linear equation as

$$\mathcal{E}(x) = \underline{\beta}'\underline{\xi}$$

and the obvious multivariate extension is

$$(\underline{x}) = B\underline{\xi}$$

where

$\underline{x}$ is an $m\times1$ vector of random variables

$\underline{\xi}$ is an $\ell\times1$ vector of error-free variables

B is an $(m\times\ell)$ matrix of (unknown) multipliers.

Our problem will be to estimate B or to perform certain tests on the nature and structure of B on the basis of a sequence of observations from the population which possesses the matrix B in the statement of its mean.

7.2 Likelihood estimation in the unrestricted parameter space (Ω).

We shall assume the sequence of observation $(\underline{x}_j)_{j=1}^{n}$ such that

$$\mathcal{E}(\underline{x}_j) = B\underline{\xi}_j \tag{7.2.1}$$

or with $X = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n)$ we shall write

$$\mathcal{E}(X) = B\Xi \tag{7.2.2}$$

where

$$\underset{(\ell \times n)}{\Xi} = (\underline{\xi}_1, \underline{\xi}_2, \ldots, \underline{\xi}_n) = (\xi_{ij}) \tag{7.2.3}$$

The matrix B is $(m \times \ell)$ and we shall assume nothing is known concerning the elements (β_{ij}).

It is to be assumed that the $\underline{x}_j$ have the same dispersion matrix and that the n vectors are independent; it is further supposed that $(\underline{x}_j)$ have multivariate normal densities. We may write therefore:

$$X \frown N_{m \times n}(B\Xi, V \otimes I). \tag{7.2.4}$$

so that

$$f(X) = \frac{|V|^{-\frac{1}{2}n}}{(2\pi)^{\frac{1}{2}mn}} \exp -\frac{1}{2} \operatorname{tr}(V^{-1}(X - B\Xi)(X - B\Xi)'). \tag{7.2.5}$$

The likelihood will be maximized with respect to B if tr $V^{-1}(X - B\Xi)(X - B\Xi)'$ is minimized with respect to B. Write

$$(7.2.6) \qquad \begin{cases} J = XX' & (m\times m) \\ H = \Xi\Xi' & (\ell\times\ell) \\ M = X\Xi' & (m\times\ell) \end{cases}$$

then we seek to minimize tr $V^{-1}(J-BM'-MB'+BHB')$ with respect to B. Using (A.4.4) and (A.4.5) of Appendix A, we find that this minimum is achieved at $\dot{B}$ given by

$$(7.2.7) \qquad \begin{cases} \dot{B}H = M \\ \dot{B}\Xi\Xi' = X\Xi' \end{cases} \qquad \text{or}$$

Now H is positive definite if the rank of Ξ is ℓ (we assume $\ell \leq n$); We assume for generality, however, that rank $\Xi = r \leq \ell$. So that the existence of H^{-1} will not be assumed.

There are $m\times\ell$ unknown elements (b_{ij}) but only $m\times r$ different equations implied by (7.2.7). We obtain a particular solution by imposing a set of $m\times(\ell-r)$ conditions

$$(7.2.8) \qquad \underset{(m\times\ell)(\ell\times\overline{\ell-r})}{\dot{B}\Delta} = \underset{(m\times\overline{\ell-r})}{(0)}$$

where, of course, Δ is chosen so that the equations (7.2.8) are not in conflict with equations (7.2.7). Any non-conflicting Δ will do, but there usually exists a Δ which has the merit of admitting simple and readily available particular solutions for B.

The particular solutions for B are obtained then from

(7.2.9) $$\begin{cases} \hat{B}H = M \\ \hat{B}\Delta = (0). \end{cases}$$

We will write

(7.2.10) $$C_\Omega = \min_B (X - B\Xi)(X - B\Xi)'$$

$$= XX' - \hat{B}H\hat{B}'$$

the "Ω" denoting minimization over the entire set (β_{ij}) and our next step is to show that C_Ω is invariant over choice of Δ. Since Ξ is $(\ell \times n)$ of rank r, there exist exactly $\ell - r$ linearly independent vectors $(\underline{k}_j)_{j=1}^{(\ell-r)}$ such that

(7.2.11) $$\underline{k}_j' \Xi = \underline{0}'.$$

Clearly if $\hat{B}_1$ and $\hat{B}_2$ are any two solutions of $\hat{B}H = M$, then corresponding rows of $\hat{B}_1$ and $\hat{B}_2$ must differ by a linear function of the form:

$$a_1\underline{k}_1' + a_2\underline{k}_2' + \dots + a_{\ell-r}\underline{k}_{\ell-r}'$$

for if not, this would imply the existence of a $\underline{k}_o'$ such that $\underline{k}_o'H = \underline{0}'$ so that $\underline{k}_o'H\underline{k}_o = (\underline{k}_o'\Xi)(\underline{k}_o'\Xi)' = 0$. Since $(\underline{k}_o'\Xi)(\underline{k}_o\Xi)'$ is a perfect square, it is implied that $\underline{k}_o'\Xi = \underline{0}'$; and since $\underline{k}_o$ is not a linear combination of the $(\underline{k}_j)_1^{\ell-r}$, the rank of Ξ must be less than or equal to $r-1$ which contradicts the definition of r.

We have then that

$$\hat{B}\Xi \text{ is invariant}$$

so that

$$\hat{B}H\hat{B}' = (\hat{B}\,\Xi)(\hat{B}\,\Xi)' \quad \text{is invariant.}$$

We see now that

$$\max_{B} f(x) = \frac{|V|^{-\frac{1}{2}V}}{(2\pi)^{\frac{1}{2}mn}}\, e^{-\frac{1}{2}\mathrm{tr}(V^{-1}C_{\Omega})} \tag{7.2.12}$$

whence $\max\limits_{B} f(x)$ is maximized over V by virtue of Theorem A.1 of Appendix A when

$$\hat{V}_{\Omega} = \frac{1}{n}\, C_{\Omega} \tag{7.2.13}$$

where the "Ω" denotes the estimate of V when B has been estimated in the unrestricted parameter space Ω. (We have used the result $(\widehat{V^{-1}} = \hat{V}^{-1})$.

We express these results in a theorem

Theorem 45

If $X \sim N_{m\times n}(B\,\Xi;\ V \otimes I)$ then the maximum likelihood estimate $\hat{B}$ of B is

$$\hat{B}_{\Omega} = (X\Xi')(\Xi\Xi')^{-1}$$

when $|\Xi\,\Xi'| \neq 0$. If $\Xi\,\Xi'$ is singular then no unique estimate $\hat{B}_{\Omega}$ exists. However, if $\hat{B}_{\Omega}$ is any solution of $\hat{B}(\Xi\,\Xi') = X\Xi'$ then

$$\hat{V}_{\Omega} = \frac{1}{n}\, C_{\Omega} = \frac{1}{n}\,(XX' - \hat{B}\Xi\,\Xi'\hat{B}')$$

is uniquely the maximum likelihood estimate of V.

7.3 Distribution of the unrestricted likelihood estimators.

Since Ξ is of rank r, it follows that we can express $(\ell-r)$ rows of Ξ in terms of the other r rows of Ξ. Alternatively, we have that there exists A, $(m \times r)$ and T, $(r \times n)$, of full rank $r<n$ such that

$$B\Xi = AT. \tag{7.3.1}$$

The representation of $B\Xi$ as AT where T is a known matrix of full rank is called a reparameterization. This reparameterization is not unique, neither is it always convenient to perform in practice; nevertheless, it can be done in theory. We may now write

$$X \frown N_{m \times n}(AT;\ V \otimes I) \tag{7.3.2}$$

and using Theorem 45

$$\left\{\begin{array}{l} \hat{A}_\Omega = (XT')(TT')^{-1} \\ \hat{V}_\Omega = \frac{1}{n}C_\Omega \\ C_\Omega = XX'-\hat{A}TT'\hat{A}'. \end{array}\right. \tag{7.3.3}$$

By Theorem 10, Chapter 2

$$\hat{A}_\Omega \frown N_{m \times r}(A;\ V \otimes (TT')^{-1}) \tag{7.3.4}$$

whence

$$\hat{B}_\Omega\Xi = \hat{A}_\Omega T \frown N_{m \times n}(B\Xi;\ V \otimes T'(TT')^{-1}T), \tag{7.3.5}$$

which implies that the calculated $T'(TT')^{-1}T$ is invariant over choice of reparameterization systems since $\hat{B}_\Omega\Xi$ is unique and therefore must have a unique distribution.

Now

$$C_\Omega = XX' - \hat{B}_\Omega\Xi\Xi'\hat{B}_\Omega'$$

$$= XX' - \hat{A}_\Omega(TT')A_\Omega'$$

$$= X(I-T'(TT')^{-1}T)X$$

and since $I-T'(TT')^{-1}T$ is idempotent of rank $n-r$, we have, by Theorem 14, Chapter 3

(7.3.8) $$\hat{C}_\Omega \frown \mathcal{W}_m(V;n-r:(0))$$

the noncentrality matrix being zero since $\mathcal{E}(x) = AT$ and

(7.3.9) $$T(I-T'(TT')^{-1}T)T'A' = (0).$$

Using theorem 16, it is readily established that $\hat{A}_\Omega$ (and therefore $\hat{A}_\Omega T = \hat{B}_\Omega\Xi$) and C_Ω have independent distributions.

We state our distributional results in two theorems.

Theorem 46 (non-singular $\Xi\Xi'$)

If $X \frown N_{m\times n}(B\Xi;\ V \otimes I)$ with Ξ an $\ell\times n$ matrix of full rank then

$\hat{B}_\Omega = X\Xi'(\Xi\Xi')^{-1}$ is the maximum likelihood estimate of B

and has the density

$$\hat{B}_\Omega \frown N_{m\times\ell}(B;\ V \otimes \Xi'(\Xi\Xi')^{-1}\Xi).$$

$\hat{V}_\Omega$ the maximum likelihood estimate of V is given by

$$\hat{V}_\Omega = \frac{1}{n}C_\Omega = \frac{1}{n}X(I-\Xi'(\Xi\Xi')^{-1}\Xi)X'$$

and $C_\Omega = n\hat{V}_\Omega \frown \mathcal{W}_m(V;n-\ell:(0))$.

The distributions of $\hat{V}_\Omega$ and $\hat{B}_\Omega$ are independent.

Theorem 47 (singular $\Xi\Xi'$)

If $X \sim N_{mxn}(B\Xi\,;\,V \otimes I)$ and Ξ is ℓxn of rank $r<\ell(<n)$, then $\hat{B}_\Omega$ is any solution of $\hat{B}_\Omega \Xi\Xi' = X\Xi'$

$\hat{B}_\Omega\Xi$ is unique and

$\hat{B}_\Omega\Xi \sim N_{mxn}(B\Xi\,;\,V \otimes \Phi)$

where Φ is the unique (nxn) idenpotent matrix of rank r satisfying $\Xi\Phi = \Xi$.

$\hat{V}_\Omega$, the maximum likelihood estimate of V, is given uniquely by

$$\hat{V}_\Omega = \frac{1}{n} C_\Omega = \frac{1}{n}(XX' - \hat{B}_\Omega \Xi\Xi' \hat{B}_\Omega')$$

and $C_\Omega = n\hat{V}_\Omega \quad {}_m(V;n-r;(0))$
the distributions of $\hat{V}_\Omega$ and $\hat{B}_\Omega\Xi$ being independent.

The Φ which occurs in the statement of this last theorem needs comment since it has not been discussed before. By comparison with equation (7.3.5) Φ is of the form $T'(TT')^{-1}T$ and is therefore idempotent (nxn) of rank r.

Comparing the densities of $\hat{B}_\Omega\Xi\,\Xi'$ and $X\Xi'$ we clearly require that

$$\Xi\Phi\Xi' = \Xi\Xi' \tag{7.3.10}$$

Normally there are many solutions for Φ of this equation (for example the nxn identity matrix) however, since we know that Φ must be nxn idenpotent of rank r, it is uniquely the idempotent rank r solution of

$$\Xi\Phi = \Xi. \tag{7.3.11}$$

7.4 Likelihood estimation in a restricted parameter space (ω). Estimable matrices.

Our model is as in section 7.2, however, now we seek to maximize the likelihood of a matrix X subject to a restriction of the form

$$BA = \Gamma_o \qquad (7.4.1)$$

where A is an ℓxs matrix of given constants and Γ_o is a given mxs matrix of constants (oftentimes a matrix of zeros in practical cases). With no loss of generality, we shall assume $BA = \Gamma_o$ actually represents a set of s equations on the elements of every row of B; a total of ms equations in all. The unknown elements of B are required to satisfy a set of ms equations and we may say that they lie in a restricted parameter space; it is convenient to write ω for this restricted parameter space and to denote the maximum likelihood estimates of B and V under these restrictions by $\hat{B}_\omega$ and $\hat{V}_\omega$ respectively.

Maximization of the likelihood with respect to B subject to $BA = \Gamma_o$ is achieved by minimizing

$$\operatorname{tr} V^{-1}(X - B\Xi\,(X - B\Xi)' \text{ subject to } BA = \Gamma_o$$

Setting up an sxm matrix of Lagrange multipliers which is conveniently taken as ΘV^{-1}, we require that

$$\left.\frac{dQ}{dB}\right|_{\hat{B}_\omega} = \left.\frac{dQ}{d\Theta}\right|_{\hat{\Theta}_\omega} = (0) \qquad (7.4.2)$$

with

(7.4.3) $$Q = \frac{1}{2}\mathrm{tr}\, V^{-1}(X-B\Xi)(X-B\Xi)' - \mathrm{tr}(BA-\Gamma_o)\Theta V^{-1}$$

this yields

(7.4.4a) $$\hat{B}_\omega \Xi\Xi' = X\Xi' + \Theta' A'$$

(7.4.4b) $$\hat{B}_\omega A = \Gamma_o$$

Let us look at some implications of equation (7.4.4a). Since there exists $\{\underline{k}_j\}_1^{\ell-r}$ such that

(7.4.5) $$\underline{k}_j' \Xi = \underline{0}'$$

we find after post multiplying (7.4.4a) by $\underline{k}_j$ that

(7.4.6) $$\underline{k}_j' A = \underline{0}' \qquad j = 1,\dots,\ell-r,$$

so that

(7.4.7) $$\mathrm{rank}\ (A) = s \leq r$$

so that the column dimension, s, of A cannot exceed the rank of Ξ. Clearly we must have

(7.4.8) $$\mathrm{rank}\ (\underset{\ell\times \overline{n+s}}{\Xi \vdots A}) = \mathrm{rank}\ (\underset{\ell\times n}{\Xi})$$

Obviously any column of A must be expressible as a linear combination of columns of Ξ. As a consequence $\hat{\Gamma}_\Omega$ defined by:

(7.4.9) $$\hat{\Gamma}_\Omega = B\ A$$

is such that $\hat{\Gamma}_\Omega$ is invariant over $\hat{B}_\Omega$ satisfying $\hat{B}_\Omega \Xi\Xi' = X\Xi'$.

Theorem 48

If A is such that $\hat{B}_\Omega A$ is invariant over any $\hat{B}_\Omega$ satisfying $\hat{B}_\Omega \Xi\Xi' = X\Xi'$ we shall describe BA as an estimable matrix and note that BA is estimable if and only if every column of A is expressible an a linear combination of columns of Ξ.

Equations (7.4.4.a), (7.4.4b) represent m(r+s) equations in m(ℓ+s) unknowns. As in section (7.2) we employ a suitable set

(7.4.10) $$\hat{B}_\omega \Delta = (0)$$

with Δ an $\ell \times \overline{\ell - r}$ matrix.

Matrix $\hat{B}_\omega$ is now the particular solution of

$$\hat{B}_\omega \Xi\Xi' = X\Xi' + \theta' A'$$

* (7.4.11) $$\hat{B}_\omega A = \Gamma_0$$

$$\hat{B}_\omega \Delta = (0)$$

and it is noted that if there exists a single solution for $\hat{B}_\omega$ then $\hat{B}_\omega \Delta$ must not be estimable, that is $(\Xi \vdots \Delta)$ must be of rank ℓ.

---

* It is easy to show that these same equations result if (X-BΞ)(X-BΞ)' is minimized with B subject to both $BA = \Gamma_0$ and $B\Delta = 0$. In practice this technique is often to be preferred since it may well emerge that some of the elements of B must be zero [See the example in section 7.6]

As before, we may show that

$\hat{B}_\omega$ is invariant over choice of Δ

whence C_ω is invariant where

(7.4.12) $$C_\omega = (X-\hat{B}_\omega\Xi)(X-\hat{B}_\omega\Xi)'.$$

The likelihood maximized over B subject to $BA = \Gamma_o$ now becomes

(7.4.13) $$\max_\omega f(x) = \frac{|V|^{-\frac{1}{2}n}}{(2\pi)^{\frac{1}{2}mn}} e^{-\frac{1}{2}\mathrm{tr}(V^{-1}C_\omega)}$$

so that, applying Theorem A.1 of appendix A,

(7.4.14) $$\hat{V}_\omega = \frac{1}{n}\hat{C} = \frac{1}{n}(X-\hat{B}_\omega\Xi)(X-\hat{B}_\omega\Xi)'$$

We express our results so far in a theorem.

Theorem 49

If $X \sim N_{m\times n}(B\Xi; V \otimes I)$ with Ξ an $m\times\ell$ matrix of rank $r\leq\ell$, then if it is known that $BA = \Gamma_o$, the likelihood of X is maximized over B by setting $B = \hat{B}_\omega$ with $\hat{B}_\omega$ any solution of

$$\hat{B}_\omega\Xi\Xi' = X\Xi'+\Theta'A'$$

$$\hat{B}_\omega A = \Gamma_o$$

$\hat{V}_\omega = \frac{1}{n}C_\omega = \frac{1}{n}(X-\hat{B}_\omega\Xi)(X-\hat{B}_\omega\Xi)'$ is uniquely the maximum likelihood estimate of V.

7.5 Distribution of the restricted likelihood estimators

As in section 7.3, we must first reparameterize the system and set

$$B\Xi = \Lambda T \quad \begin{cases} \Lambda: & (m\times r) \\ T: & (r\times n) \text{ of rank } r. \end{cases} \tag{7.5.1}$$

Incidentally, we see that

$$\Lambda = B\Xi T'(TT')^{-1} \tag{7.5.2}$$

and since $\Xi T'(TT')^{-1}$ is a matrix whose columns are linear combinations of the columns of Ξ, we see that Λ is, by our definition, an estimable matrix.

Now there exists N such that

$$A = \Xi N \qquad (N \text{ is } n\times s\text{, rank } s) \tag{7.5.3}$$

Since columns of A are linear combinations of columns of Ξ. Our restricted parameter space becomes

$$\begin{aligned} \Gamma_o &= BA \\ &= B\Xi N = \Lambda TN \end{aligned} \tag{7.5.4}$$

The solution for $\hat{\Lambda}_\omega$ is, by comparison with (7.4.11)

$$\begin{aligned} \hat{\Lambda}_\omega TT' &= XT' + \theta'N'T' \\ \hat{\Lambda}_\omega TN &= \Gamma_o \end{aligned} \tag{7.5.5}$$

these representing m(r+s) equations in the m(r+s) unknown elements of $\hat{\Lambda}$ and θ. It is convenient to write

$$\hat{\Gamma}_\Omega = \hat{\Lambda}_\Omega TN, \qquad (=\hat{B}_\Omega A) \tag{7.5.6}$$

so that

$$\hat{\Lambda}_\omega = \hat{\Lambda}_\Omega + \Theta' N' T' (TT')^{-1} \qquad (7.5.7)$$

and thence, post-multiplying by TN

$$\Gamma_o - \hat{\Gamma}_\Omega = \Theta' N' T' (TT')^{-1} TN \qquad (7.5.8)$$

Solving for Θ' and substituting back in (7.5.7)

$$\hat{\Lambda}_\omega - \hat{\Lambda}_\Omega = (\Gamma_o - \hat{\Gamma}_\Omega)[N'T'(TT')^{-1}TN]^{-1} N'T'(TT')^{-1} \qquad (7.5.9)$$

Now

$$\begin{aligned} C_\omega - C_\Omega &= (X - \hat{\Lambda}_\omega T)(X - \hat{\Lambda}_\omega T)' - (X - \hat{\Lambda}_\Omega T)(X - \hat{\Lambda}_\Omega T)' \\ &= (\hat{\Lambda}_\Omega - \hat{\Lambda}_\omega) TT' (\Lambda_\Omega - \hat{\Lambda}_\omega)' \\ &= (\hat{\Gamma}_\Omega - \Gamma_o)[N'T'(TT')^{-1}TN]^{-1}(T_\Omega - \Gamma_o)' \end{aligned} \qquad (7.5.10)$$

But

$$\begin{aligned} \hat{\Gamma}_\Omega &= XT'(T'T)^{-1}TN \\ &\frown N_{mxs}(\Gamma;\ V \otimes N'T'(TT')^{-1}TN) \end{aligned} \qquad (7.5.11)$$

where

$$\begin{aligned} \Gamma &= (XT'(T'T)^{-1}TN) \\ &= \Lambda TN \end{aligned} \qquad (7.5.12)$$

when it is supposed that

$$\mathcal{E}(X) = \Lambda T \qquad (=B\Xi). \qquad (7.5.13)$$

Rewriting $C_\omega - C_\Omega$ in terms of X we have

$$C_\omega - C_\Omega = (X-\Lambda T)(T'(TT')^{-1}TN)[N'T'(TT')^{-1}TN]^{-1} \\ (N'T'(TT')^{-1}T)(X-\Lambda T)' \tag{7.5.14}$$

The matrix of this generalized quadratic form is, by inspection, idempotent of rank s. In addition it is orthogonal to $I-T'(TT')^{-1}T$. We see then that

$$C_\omega - C_\Omega \sim \mathcal{W}_m(V;s:(\Gamma_o-\Gamma^*)(N'T''(TT')^{-1}TN)^{-1}(\Gamma_o-\Gamma^*)') \tag{7.5.15}$$

and since $N'T'(TT')^{-1}TN$ is positive definite, the noncentrality matrix of $C_\omega - C_\Omega$ is zero if and only if $\Gamma_o = \Gamma^*$. Also, by the orthogonality of the forms $C_\omega - C_\Omega$ and C_Ω, their distributions are independent (Theorem 15).

It is noted in addition that

$$\hat{\Gamma}_\Omega = \hat{B}_\Omega \Lambda \frown N_{mxs}(B_o A;\ V \otimes A'\Phi A). \tag{7.5.16}$$

We list these results in a theorem.

Theorem 50

If $X \frown N_{mxn}(B\Xi;\ V \otimes I)$ and Ξ is ℓxn of rank r, then the likelihood of X is maximized with respect to B subject to $BA=\Gamma_o$ at $\hat{B}_\omega$ which is any solution of

$$\hat{B}_\omega \Xi\Xi' = X\Xi' + \theta'A'$$

$$\hat{B}_\omega A = \Gamma_o$$

The quantity $n\hat{V}_\omega = C_\omega = (X-\hat{B}_\omega\Xi)(X-\hat{B}_\omega\Xi)$ is invariant over $\hat{B}_\omega$ and is such that

$$C_\omega - C_\Omega \sim \mathcal{W}_m\left(V; s : (\Gamma_o - \Gamma)(N'\Phi N)(\Gamma_o - \Gamma)'\right)$$

where Γ is the true value of BA

N is any solution (‡) of $\Xi N = A$

and Φ is the unique idempotent (nxn) matrix of rank r satisfying $\Xi\Phi = \Xi$.

Further $C_\omega - C_\Omega$ and C_Ω have independent distributions.

7.6 The likelihood ratio test function and its distribution.

An example: group classification.

Using the result of the previous sections, we have

$$(7.6.1) \qquad \frac{\max\limits_{\omega} f(x)}{\max\limits_{\Omega} f(x)} = \frac{\left|\frac{1}{n} C_\omega\right|^{-\frac{1}{2}n} e^{-\frac{1}{2}\mathrm{tr}(\frac{1}{n}C_\omega)^{-1}C_\omega}}{\left|\frac{1}{n} C_\Omega\right|^{-\frac{1}{2}n} e^{-\frac{1}{2}\mathrm{tr}(\frac{1}{n}C_\Omega)^{-1}C_\Omega}}$$

and since

$$(7.6.2) \qquad \mathrm{tr}(\tfrac{1}{n}C_\omega)^{-1}C_\omega = \mathrm{tr}(\tfrac{1}{n}C_\Omega)^{-1}C_\Omega = nm$$

we have

$$\frac{\max\limits_{\omega} f(x)}{\max\limits_{\Omega} f(x)} = \frac{|C_\Omega|}{|C_\omega|}^{\frac{1}{2}n}$$

small values being significant of departures from

---

‡ Although N is not uniquely determined, it is not difficult to show that $N'\Phi N$ is unique.

(7.6.3) $$H_o: \quad BA = \Gamma_o$$

We may use any monotomic function of the likelihood ratio and so take the $\frac{2}{n}^{th}$ root and obtain

(7.6.4) $$\{BA = \Gamma_o\} = \frac{|C_\Omega|}{|C_\omega|}$$

where $\{BA = \Gamma_o\}$ is to be read "the likelihood ratio test function of the null hypothesis H_o: $BA = \Gamma_o$ against general alternatives (H_1: $BA \neq \Gamma_o$). We may write

(7.6.5) $$\mathcal{L}\{BA = \Gamma_o\} = \frac{|C_\Omega|}{|C_\Omega + (C_\omega - C_\Omega)|}$$

and since (Theorems 15, 47, 50) C_Ω and $C_\omega - C_\Omega$ have independent Wishart distributions, we observe that

(7.6.6) $$\mathcal{L}\{BA = \Gamma\} \sim U'_{m:s:n-r}$$

the U-statistic becoming central if and only if H_o is true (ie B satisfies $BA = \Gamma$).

In a few cases ($m = 1, 2$; $s = 1, 2$) we have the exact percentage points of the U-statistic (H_o true). Otherwise, we will calculate the probability of obtaining a value less than the observed $\mathcal{L}\{BA = \Gamma\}$ by chance when H_o true (see equation 4.6.18). It is repeated that small values of $\mathcal{L}\{BA = \Gamma\}$ are significant of departures from H_c

The power of this test is a function of the solutions of

$$|(\Gamma_o - \Gamma)(N'\Phi N)(\Gamma_o - \Gamma)' - \theta V| = 0$$

where Γ is the true value of BA; the power is an increasing function of any latent root, the other latent roots being held fixed (see section 4.7).

It may be instructive to give an example and we choose the problem of group classification to illustrate the techniques in obtaining C_Ω and C_ω. *It should be noted very firmly, that we seldom if ever use the matrix formation set up in the previous sections of this chapter*; these matrix formulations being to ficilitate the distributional properties of C_ω and C_Ω. *Rather we we treat each design as a separate case and proceed to maximize the likelihood under restricted and unrestricted parameter conditions*.

Example: *classification by groups.*

In each of k groups a number (varying from group to group) of observations are made. Thus we observe

$$\underline{x}_{ij} \qquad j = 1, 2, \ldots, n_i; \quad k = 1, 2, \ldots, k$$

with $\underline{x}_{ij}$ an m-dimensional column vector.

We will assume the dispersion matrix of each $\underline{x}_{ij}$ is the same (V, say) and suppose the means are given by

$$\mathcal{E}(\underline{x}_{ij}) = \underline{\mu}+\underline{\alpha}_i \qquad \begin{array}{l} j = 1, \ldots, n_i \\ i = 1, \ldots, k. \end{array}$$

Thus observations within the same group have the same mean (these means may differ from group to group). We require to test the hypothesis:

H_o: $\mathcal{E}(\underline{x}_{ij})$ is constant over all observations, i.e.,

$$\underline{\mu}+\underline{\alpha}_1 = \underline{\mu}+\underline{\alpha}_2 = \underline{\mu}+\underline{\alpha}_3 = \ldots\ldots = \underline{\mu}+\underline{\alpha}_k$$

In Ω

$$f(x) = \frac{|V|^{-\frac{1}{2}N}}{(2\pi)^{\frac{1}{2}mN}} \exp -\frac{1}{2} \sum_i \sum_j (\underline{x}_{ij}-\underline{\mu}-\underline{\alpha}_i)' V^{-1} (\underline{x}_j-\underline{\mu}-\underline{\alpha}_i)$$

$$(N = \sum_i n_i)$$

this is maximized with respect to $\underline{\mu}$ and $\underline{\alpha}_i$ when

$$\sum_i \sum_j \underline{x}_{ij} = \sum_i \sum_j (\underline{\hat{\mu}}+\underline{\hat{\alpha}}_i)$$

$$\sum_j \underline{x}_{ij} = \sum_j (\underline{\hat{\mu}}+\underline{\hat{\alpha}}_i) \qquad i = 1, 2, \ldots, k.$$

We have k+1 unknown vectors but only k equations (since the first is the sum over i of the second set of equations). For a unique solution we choose a convenient equation on the $\underline{\mu}$ and $\{\alpha_i\}_1^k$ from the first equation, we could have a simple solution for $\underline{\hat{\mu}}$ and thence for $\underline{\hat{\alpha}}_i$ from the second equation. Accordingly, let us impose

$$\sum_i \sum_j \underline{\hat{\alpha}}_i = \sum_i \underline{\hat{\alpha}}_i n_i = 0$$

then our solutions are

$$\underline{\hat{\mu}} = \frac{1}{N} \sum_i \sum_j \underline{x}_{ij} = \underline{\bar{x}}_{..}$$

$$\underline{\hat{\mu}}+\underline{\hat{\alpha}}_i = \frac{1}{n_i} \sum_j \underline{x}_{ij} = \underline{\bar{x}}_{i.}$$

thus

$$\max_{\mu,\{\underline{\alpha}_i\}} f(x) = \frac{|V|^{-\frac{1}{2}N}}{(2\pi)^{\frac{1}{2}mN}} e^{-\frac{1}{2}\mathrm{tr}\, V^{-1}\sum_i\sum_j(\underline{x}_{ij}-\overline{\underline{\ }}_{i}.)(\underline{x}_{ij}-\overline{\underline{x}}_{i}.)'}$$

whence

$$N\hat{V}_\Omega = C_\Omega = \sum_i\sum_j(\underline{x}_{ij}-\overline{\underline{x}}_{i}.)(\underline{x}_{ij}-\overline{\underline{x}}_{i}.)'$$

In ω

We maximize f(x) subject to $\underline{\alpha}_1 = \underline{\alpha}_2 = \ldots\ldots = \underline{\alpha}_k$ and by virtue of the result appended as a footnote to equation (7.4.11) we minimize subject also to the condition $\sum_i n_i\underline{\alpha}_i = 0$. Clearly

$$\left.\begin{array}{l} \underline{\alpha}_1 = \underline{\alpha}_2 = \ldots\ldots = \underline{\alpha}_k \\ \sum_{i=1}^{k} \underline{\alpha}_i n_i = \underline{0} \end{array}\right\} \Longrightarrow \{\underline{\alpha}_i = \underline{0}\}_{i=1}^{k}$$

and hence the partially maximized f(x) (i.e., maximized with respect to the $\{\underline{\alpha}_i\}$ but not yet with respect to $\underline{\mu}$) becomes

$$f(x) = \frac{|V|^{-\frac{1}{2}N}}{(2\pi)^{\frac{1}{2}Nm}} e^{-\frac{1}{2}\sum_i\sum_j(\underline{x}_{ij}-\underline{\mu})'V^{-1}(\underline{x}_{ij}-\underline{\mu})}$$

Maximization with respect to $\underline{\mu}$ obtain when

$$\hat{\underline{\mu}} = \overline{\underline{x}}..$$

and then

$$N\hat{V}_\omega = C_\omega = \sum_i\sum_j(\underline{x}_{ij}-\overline{\underline{x}}..)(\underline{x}_{ij}-\overline{\underline{x}}..)'$$

These give us our analysis table

SOURCE	MATRIX	DEGREES FREEDOM
Between groups $C_\omega - C_\Omega$	$\sum_i (\overline{\underline{x}}_i - \overline{\underline{x}}_{..})(\overline{\underline{x}}_i - \overline{\underline{x}}_{..})' n_i$	k-1
Within groups C_Ω	$\sum_i \sum_j (\underline{x}_{ij} - \overline{\underline{x}}_{i.})(\underline{x}_{ij} - \overline{\underline{x}}_i)'$	N-k

$$\mathcal{L}(\underline{\alpha}_1 = \underline{\alpha}_2 = \ldots\ldots = \underline{\alpha}_k) = \frac{|C_\Omega|}{|C_\omega|} \sim U'_{m:k-1:n-k}$$

The degrees of freedom may be argued in several ways. Perhaps the easiest is as follows

C_Ω: The matrix B of the general theory is, in this case,

$$B = (\underline{\mu}, \underline{\alpha}_1, \underline{\alpha}_2, \ldots, \underline{\alpha}_k) \quad (m \times \overline{k+1})$$

so that ℓ of the general theory is here $k+1$. We needed one equation of the form $B\Delta = (0)$ to obtain a particular solution, so that in our case we actually had $\Delta = (0, n_1, n_2, \ldots, n_k)$. Since one equation was needed, the $(\overline{k+1} \times N)$ matrix Ξ, had we written it down, must have been of rank $r = k$. The degrees of freedom for C_Ω are "n-r" which becomes N-k in our case.

$C_\omega - C_\Omega$: The matrix A of the general theory becomes

$$A = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 \ldots\ldots 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \ldots\ldots 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \ldots\ldots 0 \\ & & & & & \\ 0 & 1 & 0 & 0 & 0 & 0 \ldots -1 \end{bmatrix}$$

which is $(k-1) \times (k+1)$ of full rank $s = (k-1)$ hence the degrees of freedom for $C_\omega - C_\Omega$ are $k-1$.

7.7 Hotelling's statistic in tests of linear hypothesis.

Hotelling's statistic appears frequently in tests of linear hypothesis which are simple in the exposition; this is to be anticipated since Hotelling's T^2 is a "simple" or particular case of the U-variate which is the distirbution of the U-statistic, criterion of the test of general linear hypothesis.

If we have the model

$$(7.7.1)\qquad \underline{x}_j \frown N_m(\underline{\mu};V)$$

and $\{x_j\}_1^n$ are mutually independent we may require to test the hypothesis.

$$(7.7.2)\qquad H_o:\quad \underline{\mu} = \underline{\mu}_o \quad (\underline{\mu}_o \text{ specified})$$

against a general alternative. It is simple to show that

$$(7.7.3)\qquad \left\{\frac{\max\limits_{V;\underline{\mu}=\underline{\mu}_o} f(\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n)}{\max\limits_{V,\underline{\mu}} f(\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n}\right\}^{\frac{2}{n}} = \frac{|C|}{|C+n(\overline{\underline{x}}-\underline{\mu}_o)(\overline{\underline{x}}-\underline{\mu}_o)'|}$$

$$= \frac{1}{1+n(\overline{\underline{x}}-\underline{\mu}_o)'C^{-1}(\overline{\underline{x}}-\underline{\mu}_o)}$$

The likelihood criterion is therefore a monotonic function of $n(\overline{\underline{x}}-\underline{\mu})'C^{-1}(\overline{\underline{x}}-\underline{\mu})$ where

$$(7.7.4)\qquad \sqrt{n}(\overline{\underline{x}}-\underline{\mu}_o) \frown N_m(\sqrt{n}(\underline{\mu}_o-\underline{\mu});\ V)$$

where $\underline{\mu}$ is the true expected value of each of $\underline{x}_j$ and

$$C = \sum_j (\underline{x}_j-\underline{\bar{x}})(\underline{x}_j-\underline{\bar{x}})' \sim \mathcal{W}_m(V;n-1:(0)). \qquad (7.7.5)$$

Since C and $\underline{\bar{x}}$ have independent distribution (see Theorem 16), we have (see equation 4.5.3)

$$n(\underline{\bar{x}}-\underline{\mu})'C^{-1}(\underline{\bar{x}}-\underline{\mu}) \sim T_m^2(n-1:\ \tfrac{1}{2}(\underline{\mu}_o-\underline{\mu})'V^{-1}(\underline{\mu}_o-\underline{\mu})) \qquad (7.7.6)$$

By Theorem 35

$$n(\underline{\bar{x}}-\underline{\mu}_o)'C^{-1}(\underline{\bar{x}}-\underline{\mu}_o)\ \frac{n-m}{m} \sim F_{m:n-m}(\tfrac{1}{2}(\underline{\mu}_o-\underline{\mu})'V^{-1}(\underline{\mu}_o-\underline{\mu})) \qquad (7.7.7)$$

and observe that the noncentrality parameter is strictly positive unless $\underline{\mu} = \underline{\mu}_o$.

Equation (7.7.7) enables us to construct a confidence ellipse on the unknown mean vector $\underline{\mu}$; for we may write

$$\Pr\left\{(\underline{\bar{x}}-\underline{\mu})'C^{-1}(\underline{\bar{x}}-\underline{\mu}) \le \frac{m}{n(n-m)}\ F^{(1-\alpha)}_{m:n-m}\right\} = 1-\alpha \qquad (7.7.8)$$

which the m-dimensional ellipse in the space of $\underline{\mu}$, centre $\underline{\bar{x}}$, and with equation

$$(\underline{\mu}-\underline{\bar{x}})'C^{-1}(\underline{\mu}-\underline{\bar{x}}) = \frac{m}{n(n-m)}\ F^{(1-\alpha)}_{m:n-m}$$

will contain $\underline{\mu}_o$ with probability $1-\alpha$. (That the equation is an m-dimensional ellipse follows from the fact that C is positive definite).

The two population test follows in similar manner. Given:

$$\begin{cases} \underline{x}_j \quad N_m(\underline{\mu}_x:V) \qquad j = 1,\ldots,n_x. \\ \underline{y}_j \quad N_m(\underline{\mu}_y:V) \qquad j = 1,\ldots,n_y \end{cases} \qquad (7.7.9)$$

with $\{\{x_j\}_1^{n_x} : \{y_i\}_1^{n_y}\}$ mutually independent, then we can show that the likelihood ratio test of the hypothesis $\underline{\mu}_x=\underline{\mu}_y$ (V unknown) is monotonic in

$$(7.7.10) \qquad \frac{n_x n_y}{n_x+n_y}(\bar{\underline{x}}-\bar{\underline{y}})' C_*^{-1}(\bar{\underline{x}}-\bar{\underline{y}}) \cdot \frac{n_x+n_y-m-1}{m}$$

$$\sim F_{m:n_x+n_y-m-1}\left(\tfrac{1}{2}(\underline{\mu}_x-\underline{\mu}_y)'V^{-1}(\underline{\mu}_x-\underline{\mu}_y)\right)$$

where

$$(7.7.11) \qquad C_* = \sum_{j=1}^{n_x}(\underline{x}_j-\bar{\underline{x}})(\underline{x}_j-\bar{\underline{x}})' + \sum_{j=1}^{n_y}(\underline{y}_j-\bar{\underline{y}})(\underline{y}_j-\bar{\underline{y}})'$$

Other tests of the general linear hypothesis which have an area under an F-distribution for their critical region are given in Volume II and in exercises to this chapter.

7.8 An alternative system of null hypotheses.

The reader is almost certainly familiar with the so-called 'correlated t-test'. It may be required to test whether the (eldest) daughter is, at maturity, of the same population with regard to height as her mother. We may sample a set of 2x1 vectors $\{\underline{x}_j\}_1^n$, and ask "is $\mathcal{E}(\underline{x}_j)$ of the form $\begin{bmatrix}\mu\\ \mu\end{bmatrix}$ with μ unknown?" We may state the problem in a different way as follows: If $\mathcal{E}(\underline{x}_j) = \underline{\mu} = \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}$ $j = 1,\ldots, n$, then is $\underline{\mu}$ such that $(1,-1)\underline{\mu}=0$. It will probably be remembered that in order to test this hypothesis, we construct $\{y_j = (1,-1)\underline{x}_j\}_1^n$ and test the hypothesis that the scalar quantities $\{y_i\}$ are drawn from a population with zero mean.

Clearly an immediate generalization of this problem is "given $\{\underline{x}_j\}$ independently drawn from $N_n(\underline{\mu}, V)$ is every element of $\underline{\mu}$ the same." The most general form of this type of problem is now to be discussed.

Let

$$X \sim N_{mxn}(B\Xi;\ V \otimes I) \tag{7.8.1}$$

with B, $m{\times}\ell$; Ξ, $\ell{\times}n$ of rank $r \leq \ell < n$. Elements of Ξ are known; elements of B are unknown. The dispersion matrix V is held as unknown. It is required to test a null hypothesis of the form

$$H_o: \quad KB = B_o^* \tag{7.8.2}$$

with K a specified $s{\times}m$ matrix of full rank and B_o^* an $s{\times}\ell$ matrix with specified elements. The alternative hypothesis is

$$H_1: \quad KB = B^* \quad (\neq B_o^*) \tag{7.8.3}$$

where the elements of B^* will be usually unspecified.

The test proceedure is quite logical. We construct

$$Y = KX \tag{7.8.4}$$

so that

$$\begin{cases} Y \sim N_{sxn}(B^*\Xi;\ (KVK') \times I), \text{ given } H_1 \\ Y \sim N_{sxn}(B_o^*\Xi;\ (KVK') \times I), \text{ given } H_o \end{cases} \tag{7.8.5}$$

This clearly corresponds to the type of hypothesis discussed in the earlier sections of this chapter with

$$\left\{\begin{array}{rl} s & \text{replacing } m \\ B^* & \text{replacing } B \\ \text{unknown } KVK' & \text{replacing unknown } V \\ I & \text{replacing specified } A \\ \text{specified } B^*_o & \text{replacing } \Gamma_o. \end{array}\right.$$

We will have then

$$C_\Omega = (YY' - \hat{B}^*_\Omega \Xi\Xi' \hat{B}^{*\prime}_\Omega) \tag{7.8.6}$$

where $\hat{B}^*_\Omega$ is any solution of

$$\hat{B}^*_\Omega \Xi\Xi' = Y\Xi' \tag{7.8.7}$$

and

$$C_\omega = (YY' - B^*_o \Xi\Xi' B^{*\prime}_o), \tag{7.8.8}$$

the appropriate test function is

$$\frac{|C_\Omega|}{|C_\omega|} = \mathfrak{U} \tag{7.8.9}$$

so that $\mathfrak{U} \sim \mathfrak{U}_{s:\ell:n-r}$.

It is not transparently obvious that this proceedure is a likelihood ratio proceedure, however, a little consideration is sufficient to verify that this is so.

Define an augmented matrix $\underset{(m\times m)}{K_{(a)}}$ with the specified K for its first s rows with the provision the $|K_{(a)}| \neq 0$ (possible since K is of full rank). If

(7.8.6) $$K_{(a)} = \begin{bmatrix} K \\ \cdots \\ K^* \end{bmatrix}$$

we may theoretically choose K^* to satisfy

(7.8.7) $$KVK^{*\prime} = (0)$$

so that if

$$Y_{(a)} = \begin{bmatrix} Y \\ \cdots \\ Y^* \end{bmatrix} \quad ; \quad \Gamma_{(a)} = \begin{bmatrix} \Gamma \\ \cdots \\ \Gamma^* \end{bmatrix}$$

then Y and Y^* have independent distributions for

(7.8.8) $$\begin{bmatrix} Y \\ \cdots \\ Y^* \end{bmatrix} \frown N_{m\times n}\left(\begin{bmatrix} \Gamma\Xi \\ \cdots \\ \Gamma^*\Xi \end{bmatrix} ; \begin{bmatrix} KVK' & \vdots & (0) \\ \cdots & & \cdots \\ (0) & \vdots & K^*VK^{*\prime} \end{bmatrix} \otimes I \right)$$

In Ω $\Gamma_{(a)}$ entirely unspecified

Using the previous work of this chapter $\hat{\Gamma}_{(a)\Omega}$ is any solution of

$$\hat{\Gamma}_{(a)\Omega}\Xi\Xi' = Y_{(a)}\Xi$$

and

$$C_\Omega = n(K_{(a)}VK'_{(a)})_\Omega =$$

$$\left[\begin{array}{c|c} (Y-\hat{\Gamma}_\Omega\Xi)(Y-\Gamma_\Omega\Xi)' & (0) \\ \hdashline (0) & (Y^*-\hat{\Gamma}^*_\Omega\Xi)(Y^*-\hat{\Gamma}^*_\Omega\Xi)' \end{array}\right]$$

In [illegible] $\Gamma = \Gamma_0$; Γ^* remains unspecified.

Maximization of the likelihood with respect to Γ^* yields

$$\hat{\Gamma}^*_\omega \Xi\Xi' = Y^*\Xi$$

so that

$$\hat{\Gamma}^*_\omega \Xi = \hat{\Gamma}^*_\Omega \Xi$$

$$C_\omega = n(K_{(a)}VK'_{(a)})_\omega =$$

$$\left[\begin{array}{c|c} (Y-\Gamma_o\Xi)(Y-\Gamma_o\Xi)' & (0) \\ \hline (0) & (Y^*-\hat{\Gamma}_\omega\Xi)(Y^*-\hat{\Gamma}_\omega\Xi)' \end{array}\right]$$

Since the bottom right partitions of C_Ω and C_ω are identical, it is evident that the likelihood ratio is a monotone function simply of

$$\frac{|(Y-\hat{\Gamma}_\Omega\Xi)(Y-\hat{\Gamma}_\Omega\Xi)'|}{|(Y-\hat{\Gamma}_o\Xi)(Y-\hat{\Gamma}_o\Xi)'|}$$

where $\hat{\Gamma}_\Omega$ is any solution of

$$\hat{\Gamma}_\Omega \Xi\Xi' = Y\Xi$$

$$Y = KX$$

It is noted that K^* does not enter into the test function either explicitly or implicitly; we are not troubled, therefore, by the fact that since V is unknown, it would be impossible in practice to satisfy equation (7.8.7). Some additional properties of this test situation are developed through the exercises to this chapter.

CHAPTER 8

SOME PROPERTIES OF THE LIKELIHOOD ESTIMATES

8.1 Introduction

In this chapter we look into the geometric interpretaion of the general linear hypothesis and the consequences of the form of the estimators. In a later section we look at a possible extension to the Pitman definition of closeness (1937) to the multivariate case, showing that the likelihood estimates have certain desireable properties.

8.2 The geometry of likelihood estimation.

It will be convenient to introduce a new notation. We have considered a set of (column) vectors $\{\underline{x}_j\}_1^n$ independently drawn from normal populations with common dispersion matrix and have specified that

$$(8.2.1) \qquad X = [\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n] \frown N_{mxn}(B\Xi;\ V \otimes I).$$

In this chapter we will be concerned with the row components of X rather than the column components. In the hope of reducing confusion, therefore, we introduce the notation

$$(8.2.2) \qquad X = \begin{bmatrix} \underset{\sim}{x}_1 \\ \underset{\sim}{x}_2 \\ \vdots \\ \underset{\sim}{x}_m \end{bmatrix}$$

where $\underset{\sim}{x}_\alpha$ is a row vector (1xn)

(8.2.3) $$\underset{\sim}{x}_\alpha = (x_{\alpha 1}, x_{\alpha 2}, \ldots, x_{\alpha n}).$$

We note that elements of $\underset{\sim}{x}_\alpha$ are independent, but that corresponding elements of $\underset{\sim}{x}_\alpha$, $\underset{\sim}{x}_\beta$ are correlated. In fact

(8.2.4) $$\mathcal{E}\left((\underset{\sim}{x}_\alpha - \mathcal{E}(\underset{\sim}{x}_\alpha))'(\underset{\sim}{x}_\beta - \mathcal{E}(\underset{\sim}{x}_\beta))\right) = v_{\alpha\beta} I.$$

We may regard each $\underset{\sim}{x}_\alpha$ as a random point in $\mathcal{E}_n$: the n-dimensional Euclidean space. The set of points $\{\underset{\sim}{x}_\alpha\}_1^m$, therefore, are m correlated random points in $\mathcal{E}_n$.

When we write

(8.2.5) $$\mathcal{E}(X) = \mu \quad (m\text{x}n)$$

it is implied that the average value of the random point $\underset{\sim}{x}_\alpha$ is at the point $\mu_\alpha = (\mu)_\alpha$.

Now let it be supposed that μ is of the form

(8.2.6) $$\mu = B\Xi$$

with B, m$\times\ell$ and Ξ $\ell\times m$ of rank r. Since Ξ is of rank r, there exists exactly n-r linearly independent vectors $\{\underline{f}_j\}_1^{n-r}$ such that

(8.2.7) $$\Xi\underline{f}_j = \underline{0} \quad (m\text{x}1) \quad j = 1, 2, \ldots, n-r$$

so that

(8.2.8) $$\mu\underline{f}_j = \underline{0} \quad (m\text{x}1) \quad j = 1, 2, \ldots, n-r$$

or

(8.2.9) $$\mu_\alpha\underline{f}_j = 0 \quad j = 1, 2, \ldots, n-r \; : \; \text{every } \alpha.$$

that is each μ_α lies in a subspace of $\mathcal{S}_n$ of dimension $n-(n-r) = r$. We will denote this subspace by $\mathcal{T}_r$ and call it the unrestricted space.

Suppose now we add the restriction

(8.2.10) $$BA = (0), \quad (m\times s)$$

where A is $(\ell\times s)$ of full rank s and where columns of A are linear combinations of columns of Ξ (see estimable functions, Theorem) so that we may write (compare equation 7.5.3)

(8.2.11) $$A = \Xi N \quad (N \text{ is } n\times s, \text{ rank} s)$$

then

(8.2.12) $$\boldsymbol{\mu} N = B\Xi N = BA = (0)$$

so that

(8.2.13) $$\mu_\alpha (N)_{.t} = 0 \qquad t = 1, 2, \ldots, s : \text{ every } \alpha.$$

which represents a further set of s restrictions on the elements of μ_α (the same for every α) for if not, $(N)_{.t}$, for at least one value of t, must be a linear combination of $\{f_j\}$ so that $\Xi(N)_{.t} = \underline{0}$ which implies at least one column of A is of all zeros contradicting the requirements of full rank, s, on A.

In a restricted space defined by

(8.2.14) $$\begin{cases} \boldsymbol{\mu} = B\Xi \\ BA = (0) \end{cases}$$

each $\underset{\sim}{\mu}_\alpha$ lies in a subspace of $\mathcal{J}_r$ of dimension r-s. We will denote this subspace of $\mathcal{J}_r$ by $\mathcal{H}$.

It has been seen that maximizing the likelihood with respect to location parameters is equivalent to minimizing

$$(8.2.15) \qquad Q = \operatorname{tr} V^{-1}(X-\mu)\ X-\mu)'$$

with respect to these same parameters. In Ω (unrestricted parameter space), μ must be a point of $\mathcal{J}_r$; in ω (restricted parameter space), μ must belong to the subset $\mathcal{H}$ of $\mathcal{J}_r$.

Now

$$\begin{aligned}(8.2.16) \qquad Q &= \operatorname{tr} (X-\mu)'V^{-1}(X-\mu) \\ &= \sum_{j=1}^{m} \sum_{k=1}^{m}(V^{-1})_{jk}\sum_{i=1}^{n}(x_{ji}-\mu_{ji})(x_{ki}-\mu_{ki}) \\ &= \sum_{j=1}^{m} \sum_{k=1}^{m}(V^{-1})_{jk}(\underset{\sim}{x}_j-\underset{\sim}{\mu}_j)(\underset{\sim}{x}_k-\underset{\sim}{\mu}_k)'\end{aligned}$$

and it was seen that the minimization of Q, whether $\{\underset{\sim}{\mu}_\alpha\}_1^m$ belongs to $\mathcal{J}_r$ or to $\mathcal{H}$ is the same process no matter what the value of V subject to V being kept positive definite. We would replace V by I so that $Q \to Q^*$ where

$$(8.2.17) \qquad Q^* = \sum_{j=1}^{m}(\underset{\sim}{x}_j-\underset{\sim}{\mu}_j)(\underset{\sim}{x}_j-\underset{\sim}{\mu}_j)'$$

which is the sum of the squares of the distances from the points $\underset{\sim}{x}_j$ to the points $\underset{\sim}{\mu}_j$. This sum of squares is minimized if each distance is minimized; therefore, Q^* and Q are minimized when

$\{\underset{\sim}{\mu}_j = \underset{\sim}{\ell}_j\}_{j=1}^{m}$ where $\underset{\sim}{\ell}_j$ is the foot of the perpendicular from $\underset{\sim}{x}_j$ or to the subspace $\mathcal{J}_r$ (or $\mathcal{H}$ in the restricted case). Clearly Q^* and Q have different minimal values, nevertheless, the construction of $\{\underset{\sim}{\ell}_j\}_1^m$ is the same for both Q^* and Q.

If $\bar{\mathcal{J}}_r$ is the maximal subspace of $\mathcal{S}_n$ orthogonal to $\mathcal{J}_r$ and $\{\underset{\sim}{\mu}_j\}$ are the minimizing values of $\{\underset{\sim}{\mu}_j\}$ when $\{\underset{\sim}{\mu}_j\}$ are to lie in $\mathcal{J}_r$ (unrestricted parameter case) then $(\underset{\sim}{x}_j-\underset{\sim}{\ell}_j)$ is the projection of $\underset{\sim}{x}_j$ onto $\bar{\mathcal{J}}_r$. Then $(\underset{\sim}{x}_j-\underset{\sim}{\ell}_j)(\underset{\sim}{x}_j-\underset{\sim}{\ell}_j)'$ is the square of the distance from the origin to the projection of $\underset{\sim}{x}_j$ onto $\bar{\mathcal{J}}_r$ and

$$\frac{(\underset{\sim}{x}_j-\underset{\sim}{\mu}_j)(\underset{\sim}{x}_k-\underset{\sim}{\mu}_k)'}{\sqrt{(\underset{\sim}{x}_j-\underset{\sim}{\mu}_j)(\underset{\sim}{x}_j-\underset{\sim}{\mu}_j)'(\underset{\sim}{x}_k-\underset{\sim}{\mu}_k)(\underset{\sim}{x}_k-\underset{\sim}{\mu}_k)'}}$$

is the cosine of the angle between the lines from the origin of $\mathcal{S}_n$ to the projections of $\underset{\sim}{x}_j$ and $\underset{\sim}{x}_k$ onto $\bar{\mathcal{J}}_r$. We have seen (Theorem 45; section 7.2) that

$$n\,(\hat{V}_\Omega)_{\alpha\beta} = (\underset{\sim}{x}_j-\underset{\sim}{\ell}_{j(\Omega)})(\underset{\sim}{x}_j-\underset{\sim}{\ell}_{j(\Omega)})' \tag{8.2.18}$$

where we now use the subscript Ω to denote that we are working with the space $\mathcal{J}_r$ rather than $\mathcal{H}$ so that elements of $\hat{V}_\Omega$ can be represented as distances between points in $\bar{\mathcal{J}}_r$.

8.3 The geometry of likelihood ratio tests.

The geometry of least-squares ratio test has been discussed by this writer (1964) and many of the general results in least-squares (i.e., the case m=1) go through for general m.

Suppose J_ρ is a symmetric matrix of rank ρ then $XJ X'$ will have a Wishart distribution if and only if J_ρ is idempotent. If J_ρ is idempotent then $\underset{\sim}{x} J_\rho \underset{\sim}{x}' = c^2$ is a sphere of dimension ρ (radius c) in the space $\mathcal{S}_n$ of $\underset{\sim}{x}$. As c increases from zero, the contours, $\underset{\sim}{x} J_\rho \underset{\sim}{x}' = c^2$ define a ρ-dimensional subspace of $\mathcal{S}_n$; we will call this subspace $\mathcal{J}_\rho$.

The quantity $\underset{\sim}{\mu}_\alpha J_\rho \underset{\sim}{\mu}_\alpha'$ is the α^{th} diagonal element of the noncentrality matrix of $XJ X'$ and is the square of the distance of the origin of $\mathcal{S}_n$ to the projection of $\underset{\sim}{\mu}_\alpha$ onto the space $\mathcal{J}_\rho$. The quantity

$$\frac{\underset{\sim}{\mu}_\alpha J_\rho \underset{\sim}{\mu}_\beta'}{\sqrt{(\underset{\sim}{\mu}_\alpha J_\rho \underset{\sim}{\mu}_\alpha')(\underset{\sim}{\mu}_\beta J_\rho \underset{\sim}{\mu}_\beta')}}$$

can be interpreted as the cosine of the angle between the lines from 0 to the projection of $\underset{\sim}{\mu}_\alpha$ onto $\mathcal{J}_\rho$ and $\underset{\sim}{\mu}_\beta$ onto $\mathcal{J}_\rho$.

If $\{\underset{\sim}{\mu}_\alpha\}_1^m$ belong to $\mathcal{T}_r$, then

$$(8.3.1) \qquad \underset{\sim}{\mu}_\alpha J_\rho \underset{\sim}{\mu}_\beta = 0 \qquad \text{all } \alpha \text{ and } \beta$$

if and only if the subspace $\mathcal{J}_\rho$ is orthogonal to the subspace $\mathcal{T}_r$. The maximum value of ρ corresponds to the maximal subspace of $\mathcal{S}_n$ orthogonal to $\mathcal{T}_r$ and thus $\rho \leq n-r$; this implies

Theorem 51

If $X \frown N_{m\times n}(B\Xi;\, V \otimes I)$ where Ξ is ℓxn of rank $r \leq \ell$, then there exists *uniquely* a generalized quadratic form on X which is distributed as $\mathcal{W}_m(V;n-r:(0))$; the unique quadratic form is C_Ω defined in Theorem 45. There exists no generalized quadratic form which has a central Wishart distribution with degrees of freedom greater than n-r.

By analogy, if $\underset{\sim}{\mu}_\alpha$ is constrained to lie in the subspace $\mathcal{H}$ of $\mathcal{J}_r$, then we must concern ourselves with a subspace of $\mathcal{S}_n$ orthogonal to $\mathcal{H}$; the maximal subspace of $\mathcal{S}_n$ orthogonal to $\mathcal{H}$ corresponds to the generalized quadratic form C_ω and we note that this maximal subspace contains $\mathcal{J}_{n-r}$ as a subspace. Let us denote the maximal subspace of $\mathcal{S}_n$ orthogonal to both $\mathcal{H}$ and $\mathcal{J}_{n-r}$ by $\mathcal{B}_s$, then $\mathcal{B}_s$ is the space of $C_\omega - C_\Omega$.

It has been shown that if $Y'A_1Y$ and $Y'A_2Y$ are independently distributed as Wishart variantes, then

$$A_1 \text{ and } A_2 \text{ are idempotent}$$

$$A_1A_2 = (0)$$

(see Theorem 15, section 3.3). Geometrically the subspace corresponding to $Y'A_1Y$ is orthogonal to the subspace corresponding to $Y'A_2Y$. Conversely, if the contours of $\underline{z}'A_1\underline{z} = c_1^2$ and $\underline{z}'A_2\underline{z} = c_2^2$ are systems of orthogonal spheres, then $A_1A_2 = (0)$ and the distributions on $Y'A_1Y$ and $Y'A_2Y$ are independent, Wishart distributions. Since $\mathcal{B}_s$ is orthogonal to $\mathcal{J}_{n-r}$, we uphold the already proven fact that C_Ω and $C_\omega - C_\Omega$ have independent distributions.

It is observed that the distribution $C_\omega - C_\Omega$ is central if and only if

(8.3.2) $x_\alpha B_s x'_\beta = 0$ all α and β

with B_s the matrix corresponding to the space $\mathcal{B}_s$; that is, the matrix of the generalized quadratic form which is $C_\omega - C_\Omega$. By definition of $\mathcal{B}_s$, $\{x_\alpha B_s x'_\beta\}_{\alpha=1}^{m}{}_{\beta=1}^{m}$ will be zero whenever $\{x_\alpha\}_1^m$ is a point of $\mathcal{H}$. If, however, x_α lies outside $\mathcal{H}$ (but still is $\mathcal{J}_r$ of course) then $x_\alpha B_s x'_\alpha > 0$. We state our new results in a theorem.

Theorem 52

If $X \frown N_{mxn}(B\Xi;\, V \otimes I)$ with Ξ ℓxn of rank r and if under H_o: $BA = (0)$ where A is (ℓxs) of full rank s ($\leq \ell$), then there exists uniquely a generalized quadratic form on X

(i) which has a distribution independent of the unique C_Ω (defined in Theorem 51)

(ii) which has a Wishart distribution with s degrees of freedom

(iii) which has noncentrality matrix of (0) if and only if H_o is true.

This unique generalized quadratic form is $C_\omega - C_\Omega$ where C_ω is defined in Theorem 49.

There are some problems left to be answered. Suppose

(8.3.3) $E \frown \mathcal{W}_m(V;e:(0))$

and

$$H \sim \mathcal{W}_m(V;h:\Lambda) \tag{8.3.4}$$

such that H and E are independently distributed. If we construct

$$U = \frac{|E|}{|E+H|} \tag{8.3.5}$$

how does $\Pr[u \leq t]$ behave if we

i increase e, h and Λ held fixed?
ii decrease h, e and Λ held fixed?
iii change Λ, e and h held fixed?

These questions are answered when $m = 1$, but a full investigation of the properties of U for general m is incomplete.

8.4 The accuracy of a vector of normally distributed estimators.

A comparison of two estimators for an unknown (single) parameter is almost always effected by consideration of their variances. Another method of comparison was proposed by Pitman (1937), but has remained comparatively undeveloped due to its intractibility in many situations. In essence Pitman proposes that if t_1 and t_2 be unbiassed for some parameter (μ, say), then t_1 shall be "better" than t_2 if

$$\Pr[|t_1-\mu|<|t_2-\mu|] > \frac{1}{2}$$

In this section we offer an extension to this definition in the following way: Let

$$X \sim N_{m\times n}(B\Xi;\ V \otimes I) \tag{8.4.1}$$

with the usual properties on B, Ξ and V. We shall suppose

(8.4.2) $$\underline{\mu} = B\underline{\nu}$$

to be an estimable function (so that $\underline{\nu}$ must be a linear combination of columns of Ξ). Let $\underline{\hat{\mu}}_{(1)}$ and $\underline{\hat{\mu}}_{(2)}$ be alternative estimators for $\underline{\mu}$, then we shall say that $\underline{\hat{\mu}}_{(1)}$ is more accurate than $\underline{\hat{\mu}}_{(2)}$ if for every positive definite or semi-definite matrix A

(8.4.3) $$\Pr\left((\underline{\hat{\mu}}_{(1)}-\underline{\mu})'A(\underline{\hat{\mu}}_{(1)}-\underline{\mu})<(\underline{\hat{\mu}}_{(2)}-\underline{\mu})'A(\underline{\hat{\mu}}_{(2)}-\underline{\mu})\right) > \frac{1}{2}$$

An estimator $\underline{\hat{\mu}}_{(*)}$ will be deemed "most accurate" if $\underline{\hat{\mu}}_{(*)}$ is more accurate than any other estimator.

These definition include Pitman's definition as the case $m = 1$. The definition is attractive since for non-negative A the expression $(\underline{\hat{\mu}}-\underline{\mu})'A(\underline{\hat{\mu}}-\underline{\mu})$ will be a weighted measure of the distance of $\underline{\hat{\mu}}$ from $\underline{\mu}$; intuitively, we require of our estimators that such distances be as small as possible. Crudely stated, our definition says that no matter how we weigh distances (provided we give them a positive sign) $\underline{\hat{\mu}}_{(1)}$ will lie closer to $\underline{\mu}$ than will $\underline{\hat{\mu}}_{(2)}$ more than 50% of the time.

Let

(8.4.4) $$\underline{\hat{\mu}}_{(t)} = X\underline{\lambda}_{(t)} \qquad t = 1, 2, \ldots$$

be an estimate of $\underline{\mu}$. If $\underline{\hat{\mu}}_{(t)}$ is unbiassed, then

(8.4.5) $$\Xi\underline{\lambda}_{(t)} = \underline{\nu} \qquad t = 1, 2, \ldots$$

Consider two different estimators ($t = 1, 2$) then

(8.4.6) $$\begin{bmatrix} \underline{\beta}_{(1)} \\ \cdots\cdots \\ \underline{\beta}_{(2)} \end{bmatrix} \frown N_{2m}\left(\begin{bmatrix} \underline{\mu} \\ \cdots \\ \underline{\mu} \end{bmatrix} ; \begin{bmatrix} \theta_{11}V & \vdots\,\theta_{12}V \\ \cdots\cdots & \cdots\cdots \\ \theta_{21}V & \vdots\,\theta_{22}V \end{bmatrix} \right)$$

where

(8.4.7) $$\theta_{ts} = \theta_{st} = \underline{\lambda}'_{(t)}\underline{\lambda}_{(s)} \qquad \begin{matrix} s \\ t \end{matrix} = 1, 2$$

if

(8.4.8) $$P_{1:2}(A,\nu,V) =$$

$$\Pr\Big((\underline{\beta}_{(1)}-\underline{\mu})'A(\underline{\beta}_{(1)}-\underline{\mu}) \leq (\underline{\beta}_{(2)}-\underline{\mu})'A(\underline{\beta}_{(2)}-\underline{\mu})\Big)$$

for non-negative A then, writing

(8.4.9) $$\underline{y}_{(t)} = M(\underline{\beta}_{(t)}-\underline{\mu}) \qquad t = 1, 2$$

where M is such that

(8.4.10) $$MVM' = I$$

we have

(8.4.11) $$P_{1:2}(A,\nu,V) = \Pr\left(\underline{y}'_{w}N\underline{y}_{(1)} \leq \underline{y}'_{(2)}N\underline{y}_{(2)}\right)$$

where

(8.4.12) $$N = (M^{-1})'AM^{-1}$$

and is non-negative.

Now

$$\begin{bmatrix} \underline{y}_{(1)} \\ \cdots \\ \underline{y}_{(2)} \end{bmatrix} \frown N_{2m}\left(\begin{bmatrix} \underline{0} \\ \cdots \\ \underline{0} \end{bmatrix} ; \begin{bmatrix} \theta_{11}I & \vdots & \theta_{12}I \\ \cdots & \cdots & \cdots \\ \theta_{21}I & \vdots & \theta_{22}I \end{bmatrix}\right) \tag{8.4.13}$$

so that, making the transformation

$$\begin{cases} \underline{w}_{(1)} = (\theta_{22}\underline{y}_{(1)} - \theta_{12}\underline{y}_{(2)}) \Big/ [\theta_{22}(\theta_{11}\theta_{22}-\theta_{12}^{2})]^{\frac{1}{2}} \\ \\ \underline{w}_{(2)} = \underline{y}_{(2)} \Big/ \theta_{22}^{\frac{1}{2}} \end{cases} \tag{8.4.14}$$

we find

$$P_{1:2}(A,\nu,V) = \tag{8.4.15}$$

$$\Pr\{(\underline{W}_{(1)}+\alpha\underline{W}_{(2)})'N(\underline{W}_{(1)}-\beta\underline{W}_{(2)}) \le 0\}$$

where

$$\begin{cases} \alpha = \dfrac{(\theta_{22}+\theta_{12})}{(\theta_{11}\theta_{22}-\theta_{12}^{2})^{\frac{1}{2}}} \\ \\ \beta = \dfrac{(\theta_{22}-\theta_{12})}{(\theta_{11}\theta_{22}-\theta_{12}^{2})^{\frac{1}{2}}} \end{cases} \tag{8.4.16}$$

and

$$\begin{bmatrix} \underline{w}_{(1)} \\ \cdots\cdots \\ \underline{w}_{(2)} \end{bmatrix} \frown N_{2m}\left(\begin{bmatrix} \underline{0} \\ \cdots \\ \underline{0} \end{bmatrix} ; \begin{bmatrix} I & \vdots & (0) \\ \cdots & \vdots & \cdots \\ (0) & \vdots & I \end{bmatrix}\right) \tag{8.4.17}$$

The characteristic function of the argument inside the probability statement of (8.4.15) is

(8.4.18) $$\phi(s) = \mathcal{E}\exp\{is\,(\underline{w}_{(1)}+\alpha\underline{w}_{(2)})'N(\underline{w}_{(1)}-\beta\underline{w}_{(2)})\}$$
$$= |I+\varepsilon_1 s_i N|^{-\frac{1}{2}}|I-\varepsilon_2 s_i N|^{-\frac{1}{2}}$$

where

(8.4.19)
$$\varepsilon_1 = \frac{\sqrt{(1+\alpha^2)(1+\beta^2) + (\alpha\beta-1)}}{(\alpha+\beta)^2}$$
$$\varepsilon_2 = \frac{\sqrt{(1+\alpha^2)(1+\beta^2) - (\alpha\beta-1)}}{(\alpha+\beta)^2}$$

the sign of the surd being positive in both cases.

This characteristic function is the characteristic function of the scalar variable:

(8.4.20) $$d = \varepsilon_2\underline{z}'_{(2)}N\underline{z}_{(2)} - \varepsilon_1\underline{z}'_{(1)}N\underline{z}_{(1)}$$

where $\underline{z}_{(1)}$ and $\underline{z}_{(2)}$ are independent (mxl) vectors of independent standard normal variates; hence

(8.4.21) $$P_{1:2}(A,\nu,V) = \Pr\,\{\varepsilon_2\underline{z}'_{(2)}N\underline{z}_{(2)}<\varepsilon_1\underline{z}'_{(1)}N\underline{z}_{(1)}\}$$

But $\underline{z}'_{(2)}N\underline{z}_{(2)}$ and $\underline{z}'_{(1)}N\underline{z}_{(1)}$ have identical distributions so that, if N is non-negative

(8.4.22) $$P_{1:2}(A,\nu,V) > \frac{1}{2} \quad \text{if and only if } \varepsilon_2<\varepsilon_1.$$

Since

(8.4.23) $$\varepsilon_1-\varepsilon_2 = (\theta_{22}-\theta_{11})\Big/2\theta_{22}$$

(8.4.24) $$p_{1:2}(A,\nu,V) > \frac{1}{2} \text{ if and only if } \theta_{11}<\theta_{22}.$$

Evidently the most accurate estimator is

(8.4.25) $$\underline{\mu}_{(*)} = X\underline{\lambda}_{(*)}$$

where $\underline{\lambda}_{(*)}$ is the vector which minimized $\underline{\lambda}'\underline{\lambda}$ subject to $\Xi\underline{\lambda} = \underline{\nu}$. It is now easy to show that

(8.4.26) $$\underline{\mu}_{(*)} = \hat{B}_{\Omega}\underline{\nu} \quad \text{(uniquely)}$$

where $\hat{B}_{\Omega}$ is any solution of

(8.4.27) $$\hat{B}_{\Omega}\Xi\Xi' = X\Xi.$$

Our result is then

Theorem 53

If $X \frown N_{mxn}(B\Xi;\ V \otimes I)$ and $\underline{\mu} = B\underline{\nu}$ is estimable then

$$\underline{\mu}_{(*)} = \hat{B}_{\Omega}\underline{\nu}$$

where $\hat{B}_{\Omega}$ is any solution of

$$\hat{B}_{\Omega}\Xi\Xi' = X\Xi$$

is most accurate amongst estimators which are linear combinations of the columns of X.

The geometric aspects of this result seem to be closely linked with the geometric aspects discussed earlier in this chapter. It may well be that an approach of this type would be useful in answering the question posed at the end of the last section.

CHAPTER 9

TEST OF HYPOTHESIS CONCERNING THE FORM OF THE DISPERSION MATRIX

9.1 Introduction

In the univariate case we are concerned with a single variance (σ^2) rather than a matrix of variances and covariances (V). The number and nature of tests on a single parameter are obviously very limited. Given a single population about the only test we may meaningfully make is that relevant to a hypothesis $\sigma=\sigma_o$. Given a sequence of populations with variances σ_1^2, σ_2^2, $\sigma_3^2,\ldots$, we may perform a test of homoscedasticity, the null hypothesis starting $\sigma_1 = \sigma_2 = \sigma_3 = \ldots$. The nature of the possible null hypotheses can obviously be more varied for general m; for example:

(i) Is $V = V_o$? (V_o fully specified)

(ii) Does a certain set of elements of V takes on a corresponding set of specified values?

(iii) Given a sequence of populations with dispersion matrices $V_1, V_2, V_3,\ldots$, is $V_1=V_2=V_3=\ldots$?

It is not possible to deal with these questions under a general hypothesis as was the case with the hypotheses on location parameters. Rather each case is treated on its merits by a likelihood ratio approach, the distribution theory and techniques of earlier chapters giving us a springboard from which to establish an appropriate critical region. We shall be brief in this chapter with regard to actual derivations of the likelihood ratios, often merely stating them.

9.2 To test that the dispersion matrix is equal to a fully specified; H_o: $V = V_o$.

Given $X \frown N_{mxn}(B\Xi;\ V \otimes I)$; $\Xi(\ell xn)$ rank r.

In Ω B and V unknown

(9.2.1) $\hat{B}_\Omega$ any solution of $\hat{B}_\Omega \Xi\Xi' = X\Xi'$

(9.2.2) $n\hat{V}_\Omega = C_\Omega = (X-\hat{B}_\Omega\Xi)(X-\hat{B}_\Omega\Xi)'$

In ω B unknown; $V = V_o$ entirely specified

(9.2.3) $\hat{P}_\omega$ any solution of $\hat{B}_w \Xi\Xi' = X\Xi'$

(9.2.4) $\hat{V}_\omega = V_o$

The likelihood ratio is:

(9.2.5) $$\lambda = \frac{e}{n}^{\frac{1}{2}mn} |V_o^{-1}C_\Omega|^{\frac{1}{2}n} e^{-\frac{1}{2}\mathrm{tr}\ V_o^{-1}C}.$$

The distribution of C_Ω is $\mathcal{W}_m(V;n-r:(0))$ hence

(9.2.6) $$\mathcal{E}(\lambda^h) = \frac{2e}{n}^{\frac{1}{2}mnh} \frac{|V_o|^{\frac{1}{2}\nu}|V|^{\frac{1}{2}nh}}{|V_o+hV|^{\frac{1}{2}(\nu+nh)}} \prod_{j=1}^{m} \frac{\Gamma\left(\frac{\nu+nh+1-j}{2}\right)}{\Gamma\left(\frac{\nu+1-j}{2}\right)}$$

$(\nu = n-r)$

(9.2.7)
$$\mathcal{E}(e^{-2\theta\log\lambda}|\,H_o)$$
$$= \left(\frac{2e}{n}\right)^{-mn\theta} \frac{1}{(1-2\theta)^{\frac{1}{2}m(\nu-2\theta n)}} \prod_{j=1}^{m} \frac{\Gamma\left(\frac{\nu+1-j}{2} - \theta n\right)}{\Gamma\left(\frac{\nu+1-j}{2}\right)}$$
$$= \prod_{j=1}^{m} \left(\frac{2e}{n}\right)^{-\theta n} \frac{1}{(1-2\theta)^{\frac{1}{2}(\nu-2\theta n)}} \frac{\Gamma\left(\frac{\nu+1-j}{2} - \theta n\right)}{\Gamma\left(\frac{\nu+1-j}{2}\right)}$$

expanding the Γ-functions

(9.2.8)
$$(e^{-2\theta\log\lambda}|H_o) = \prod_{j=1}^{m} \frac{1}{(1-2\theta)^{\frac{1}{2}j}}$$

The critical region for a test based on -2logλ is therefore approximately the critical region of a $\chi^2_{\frac{1}{2}m(m+1)}$ variate; small values of λ and therefore large values of -2logλ are significant of departures from H_o.

It would be possible to improve upon the accuracy of the critical region by considering the expansion of the Γ-function ratio (similar to section 4). The $\chi^2_{\frac{1}{2}m(m+1)}$ approximation, however, is very adequate whenever n-r-m is not too small. The number of times when V is entirely specified, however, must be small indeed so that further work in this particular case is likely to be of little application.

9.3 To test that a dispersion matrix is proportional to a specified matrix; H_o : $V \propto V_o$.

Given $X \sim N_{mxn}(B\Xi;\ V \otimes I)$ with Ξ, (ℓxn) of rank r.

In Ω B and V unknown

(9.3.1) $\hat{B}_\Omega$ any solution of $\hat{B}_\Omega \Xi\Xi' = X\Xi'$

(9.3.2) $n\hat{V}_\Omega = C_\Omega = (X - \hat{B}_\Omega \Xi)(X - \hat{B}_\Omega \Xi)'$

In ω B unknown; $V = \sigma^2 V_o$; V_o given, σ^2 unknown

(9.3.3) $\hat{B}_\omega$ any solution of $\hat{B}_\omega \Xi\Xi' = X\Xi'$

(9.3.4) $$^2 = \frac{1}{nm} \operatorname{tr} V_o^{-1} C_\Omega$$

(9.3.5) $$\hat{V}_\omega = \hat{\sigma}^2 V_o.$$

The likelihood ratio is

(9.3.6) $$\lambda = \frac{|C_\Omega|^{\frac{1}{2}n}}{|V_o(\frac{1}{m}\operatorname{tr} V_o^{-1}C_\Omega)|^{\frac{1}{2}n}} = \frac{|V_o^{-1}C_\Omega|^{\frac{1}{2}n}}{(\frac{1}{n}\operatorname{tr} V_o^{-1}C_\Omega)^{\frac{1}{2}nm}}$$

It is interesting to give an interpretation to this criterion by observing that $|V_o^{-1}C_\Omega|$ is the product of the latent roots of C_Ω in the metric of V_o whereas $\operatorname{tr}(V_o^{-1}C_\Omega)$ is the sum of the latent roots of C_Ω in the metric of V_o. Denote the latent roots by $\delta_1, \delta_2, \dots, \delta_m$ then

(9.3.7) $$\frac{|V_o C_\Omega|^{\frac{1}{m}}}{(\frac{1}{m}\operatorname{tr}(V_o^{-1}C_\Omega)} = \frac{\sqrt[m]{\delta_1\delta_2\cdots\delta_m}}{\frac{1}{m}(\delta_1+\delta_2+\dots+\delta_m)}$$

which is the ratio of the geometric to arithmetic means of the m latent roots and is a reasonable measure of how there latent roots differ.

Writing

(9.3.8) $$\eta = \lambda^{\frac{2}{n}}$$

we have, under H_o

(9.3.9) $$\eta | H_o \sim \frac{|C|}{(\frac{1}{m}\mathrm{tr}(C))^m}$$

where

(9.3.10) $$C \sim \mathcal{W}_m(I; \nu = n-r : (0))$$

since we may transform using

(9.3.11) $$C = MC_{\Omega}M' \qquad \text{with } \sigma^2 MV_oM' = I$$

Now

(9.3.12) $$\mathcal{E}(\eta)^h = \int\cdots\int_{\text{pos.def.C.}} \frac{K_n(I,\nu)|C|^{\frac{\nu-m+1}{2}+h} e^{-\frac{1}{2}\mathrm{tr}(c)}}{(\frac{1}{m}\mathrm{tr}(c))^{mh}}$$

where $K_n(I,\nu)$ is the constant part of the density $\mathcal{W}_m(I;\nu:(0))$. We may think of this as

(9.3.13) $$\mathcal{E}(\eta)^h = \int\cdots\int_{\text{pos.def.C.}} \frac{K_m(I,\nu)}{K_m(I,\nu+2h)} \frac{1}{(\frac{1}{m}\mathrm{tr}(c))^{mh}} p(C^*)$$

where $p(C^*)$ is the density of $C^* \sim \mathcal{W}_m(I,\nu+2h:(0))$.

Since it is established that $\mathrm{tr}(C^*) \sim \chi^2_{(\nu+2h)m}$ (see Theorem 27, section 3.8)

(9.3.14) $$\mathcal{E}(\eta)^h \equiv m^{mh}\frac{K_m(I,\nu)}{K_m(I,\nu+2h)} \cdot \mathcal{E}\left\{\chi^2_{(\nu+2h|m)}\right\}^{-mh}$$

Simplifying the right hand side and noting that the negative moment does indeed exist, we have

(9.3.15) $$\mathcal{E}(\eta)^h = m^{mh}\frac{\Gamma(\frac{1}{2}m\nu)}{\Gamma(\frac{1}{2}m(\nu+2h))}\prod_{j=1}^{m}\frac{\Gamma\left(\frac{1}{2}(\nu+1-j)+h\right)}{\Gamma\left(\frac{1}{2}(\nu+1-j)\right)}$$

$(\nu = n-r)$

Special case: m=2 -

$$\mathcal{E}(\eta^h|H_o: \ m=2) = \frac{\nu-1}{\nu+2h-1} \qquad (\nu = n-r)$$

which is the hth moment of the density

$$p(\eta) = \tfrac{1}{2}(\nu-1)\eta^{\frac{1}{2}(\nu-3)}$$

so that we have the density of η when m=2.

Clearly

$$p(\eta \leq t) = t^{\frac{1}{2}(\nu-1)}$$

which gives very simply, the percentage points. (small values of η are significant).

The general case: $m \geq 3$

We use the approach developed by Box and illustrated in the derivation of the asymptotic distribution of the logarithm of a U-statistic.

With

$$\eta = \lambda^{\frac{1}{2}n} \quad \text{and} \quad \nu = n-r$$

$$\rho = 1 - \frac{2m^2+m+2}{6m\nu}$$

$$\gamma = \frac{(m-1)(m^2-4)(2m^3+6m^2+3m+2)}{288(m\nu\rho)^2}$$

$$f = \tfrac{1}{2}m(m+1) - 1$$

then

$$\Pr\left(-\rho\nu\log\eta \leq t\right) = \Pr\left(-2\frac{\rho\nu}{n}\log\lambda \leq t\right)$$

$$= \Pr\{\chi^2_f \leq t\} + \gamma[\Pr\{\chi^2_{f+4} \leq t\} - \Pr\{\chi^2_f \leq t\}]$$

$$+ \text{ error of order } \nu^{-3}.$$

[Note that the result obtained by ignoring the error term becomes exact if m is set equal to 2.]

9.4 To test that a number of dispersion matrices are equal (test of homoscedasticity).

Given a sequence of k populations $\Pi_1, \Pi_2, \ldots, \Pi_k$ suppose Ξ_t has rank r_t and

(9.4.1) $$X_t \frown N_{m\times n_t}(B_t\Xi_t;\ V_t \otimes I) \qquad t = 1, 2, \ldots, k$$

observations from different populations are independent.

In Ω B_t, V_t (assumed different) all unknown.

(9.4.2) $\hat{B}_t$ is any solution of $\hat{B}_{t\Omega}\Xi\Xi' = X_t\Xi$

(9.4.3) $$n_t\hat{V}_{t\Omega} = C_{t\Omega} = (X_t-\hat{B}_{t\Omega}\Xi_t)(X_t-\hat{B}_{t\Omega}\Xi_t)'$$

for each t.

In ω $\{B_t\}$ unknown $V_1=V_2=\ldots=V_k=V$ say, unknown estimates of location are as in Ω.

(9.4.4) $$N\hat{V}_\omega = \sum_{t=1}^{k} C_{t\Omega} \qquad (N = \sum_{1}^{k} n_t)$$

The likelihood ratio is

(9.4.5) $$\lambda = \frac{\prod_{t=1}^{k} \left| \frac{1}{n_t} C_{t\Omega} \right|^{\frac{1}{2}n_t}}{\left| \frac{1}{N} \sum_{t=1}^{k} C_{t\Omega} \right|^{\frac{1}{2}N}}$$

Defining

(9.4.6) $$\eta^{\frac{1}{2}N} = \frac{\prod_{t=1}^{k} \left| C_{t\Omega} \right|^{\frac{1}{2}n_t}}{\left| \sum_{1}^{k} C_{t\Omega} \right|^{\frac{1}{2}N}}$$

Writing $\alpha_t = n_t/_N$ (so that $\sum_{1}^{k} \alpha_t = 1$)

we have

(9.4.7) $$\eta = \prod_{t=1}^{k} \left| C_{t\Omega} \right|^{\alpha_t} \Big/ \left| \sum C_{t\Omega} \right|$$

If H_o is true we may transform

(9.4.8) $$C_t = MC_{t\Omega}M'$$

using the M for which MVM' = I. Thus

(9.4.9) $$\eta | H_o \sim \prod_{t=1}^{k} \left| C_t \right|^{\alpha_t} \Big/ \left| \sum C_t \right|$$

where

(9.4.10) $$C_t \frown \mathcal{W}_m(I;\nu_t:(0)) \qquad t = 1, 2, \ldots, k \quad (\nu_t = n_t - r_t)$$

and the $\{C_t\}_1^k$ are mutually independent. We seek the h^{th} moment of η under H_o.

(9.4.11) $$\mathcal{E}(\eta^h|H_o) = \underset{\{\text{pos.def.}C_t\}_1^k}{\int\cdots\int} \frac{\prod_{t=1}^{k} K(I;\nu_t)|C_t|^{\frac{1}{2}(\nu_t-m+1)+h\alpha_t} e^{-\frac{1}{2}tr(C_t)}}{|\sum C_t|^h}$$

where $K(I;\nu_t)$ is the constant part of the density $\mathcal{W}_m(I;\nu_t:(0))$ that is

(9.4.12) $$K(I;\nu_t) = \frac{1}{2^{\frac{1}{2}m\nu_t}\pi^{\frac{1}{4}m(m-1)}\prod_{j=1}^{m}\Gamma(\frac{1}{2}(\nu_t+1-j))}$$

so that

(9.4.13) $$\mathcal{E}(\eta^h|H_o) = \prod_{t=1}^{k}\frac{K(I;\nu_t)}{K(I;\nu_t+\alpha_t h)} \underset{\{C_t^*\}}{\mathcal{E}} |\sum C_t^*|^{-h}$$

where $\{C_t^* \frown \mathcal{W}_m(I;\nu_t+\alpha_t h:(0))\}_1^k$ and are independent. Now

(9.4.14) $$\sum C_t^* \frown \mathcal{W}_m(I;\sum_1^k \nu_t + h:(0))$$

(see Theorem 12, section 3.2 and last paragraph of section 5.4).

Finally.

$$(9.4.15) \qquad \mathcal{E}(\eta^h|H_0) = \frac{K(I;h+\sum_1^k \nu_t)}{K(I;\sum_1^k \nu_t)} \left\{ \prod_{t=1}^{k} \frac{K(I;\nu_t)}{K(I;\nu_t+\alpha_t h)} \right\}$$

which can be expressed in terms of the Γ-functions implicit in the $K(I;\nu_t)$ etcetera. Using Box's technique we could give an asymptotic approximation for the distribution of $-2\rho\log\eta$ for a suitably chosen ρ. In actual test of hypothesis, it has become customary to consider a slightly different statistic wherein n_t is replaced by ν_t (the degrees of freedom associated with $C_{\Omega t}$.

We will write

$$(9.4.16) \qquad u = \frac{(\sum\nu_t)^{\frac{1}{2}m\sum\nu_t}}{\prod_{t=1}^{k}\nu_t^{\frac{1}{2}m\nu_t}} \left\{ \frac{\prod_{t=1}^{k} |C_{t\Omega}|^{\beta_t}}{|\sum_{t=1}^{k} C_{t\Omega}|} \right\}^{\frac{1}{2}\sum\nu_t} \quad \left(\beta_t = \frac{\nu_t}{\sum\nu_t}\right)$$

the term within the bracket ({...}) has the same moment as has η except that α_t must be replaced by $\beta_t = \nu_t / \sum\nu_t$. We now apply Box's technique to u and obtain

$$\Pr\{-2\rho\log u \le t\} = \Pr\{\chi_f^2 \le t\} + \gamma[\Pr\{\chi_{f+4}^2 \le t\} - \Pr\{\chi_f^2 \le t\}] + \text{error of order } (\sum\nu_t)^{-3}$$

when

$$\rho = 1 - \frac{2m^2+3m-1}{6(m+1)(k-1)}\left(\sum\frac{1}{\nu_t} - \frac{1}{(\sum\nu_t)}\right)$$

$$48\rho^2\gamma = m(m^2-1)(m+2)\left(\sum\frac{1}{\nu_t^2} - \frac{1}{(\sum\nu_t)^2}\right) - 6m(m+1)(k-1)(1-\rho)^2$$

$$f = \frac{1}{2}m(m+1)(k-1)$$

A further alternative is suggested which is to replace $-2\rho\log u$ by $(-2\log u)/(2-\rho)$. Since $\frac{1}{2-\rho} = 1-(1-\rho)+(1-\rho)^2-\ldots+\ldots$etc., there will be little difference when ρ is close to unity (i.e., when the ν_t are large).

It is noted that the univariate test statistic in the univariate case can be obtained by replacing m=1 when we get, using the criterion $(-2\log u)/(2-\rho)$,

$$(\sum\nu_t)\log_e\left(\frac{\sum s_t^2}{\sum\nu_t}\right) - \sum_{t=1}^{k}\nu_t\log_e\left(\frac{s_t^2}{\nu_t}\right)$$

$$\overset{\cdot}{\underset{\cdot}{\frown}} \quad \left(1+\frac{3}{k-1}\left(\sum\frac{1}{\nu_t} - \frac{1}{\sum\nu_t}\right)\right) \quad \chi^2_{k-1}$$

with $\frac{1}{\nu_t} s_t^2$ an unbiassed estimate of σ_t^2 (the variance in the tth population) an $s_t^2 \quad \sigma_t^2\chi^2_{\nu_t}$.

9.5 To test for independence between two groups of observations.

Suppose $X \frown N_{m\times n}(B\Xi;\ V \otimes I)$ with Ξ of rank r. It is required to test H_o: $V_{12} = (0)$ where V_{12} is say the upper-right pxq partition of V ($p+q = m$; $V_{12} = V'_{21}$).

In Ω B and V entirely unknown.

(9.5.1) $\hat{B}_\Omega$ any solution of $\hat{B}_\Omega\Xi\Xi' = X\Xi'$

(9.5.2) $n\hat{V}_\Omega = C_\Omega = (X - \hat{B}_\Omega\Xi)(X - \hat{B}_\Omega\Xi)'$

In ω B unknown; $V = \left[\begin{array}{c:c} V_{11} & (0) \\ \hdashline (0) & V_{22} \end{array}\right]$, V_{11}, V_{22} unknown.

(9.5.3) $\hat{B}_\omega$ is any solution of $\hat{B}_\omega \Xi\Xi' = X\Xi'$

$$nV_\omega = \left[\begin{array}{c:c} C_{\Omega 11} & (0) \\ \hdashline (0) & C_{\Omega 22} \end{array}\right] \tag{9.5.4}$$

that is $\hat{V}_\omega$ is simply obtained by replacing the off-diagonal partitions of C_Ω (corresponding to V_{12} and V_{21}) by matrices of zeros.

The likelihood ratio λ is such that

$$\lambda^{2/n} = \frac{|C_\Omega|}{|C_{\Omega 11}||C_{\Omega 22}|} = \eta \text{ say} \tag{9.5.5}$$

if H_o be true, then by an appropriate transformation

$$C = \left[\begin{array}{c:c} K_{11} & (0) \\ \hdashline (0) & K_{22} \end{array}\right] C_\Omega \left[\begin{array}{c:c} K_{11} & (0) \\ \hdashline (0) & K_{22} \end{array}\right]' \tag{9.5.6}$$

where

$$\left\{\begin{array}{l} K_{11}V_{11}K_{11}' = I \\ \\ K_{22}V_{22}K_{22}' = I \end{array}\right. \tag{9.5.7}$$

then

$$\eta | H_o \sim |C| \Big/ |C_{11}||C_{22}| \tag{9.5.8}$$

where C_{11} is the top-left pxp partition of C and C_{22} is the bottom-right qxq partition of C.

Now

$$C \sim \mathcal{W}_m(I;\nu:(0)) \qquad (\nu=n-r) \tag{9.5.9}$$

and using the clockwise rule

$$\eta|H_o \sim \frac{|C_{11}-C_{12}C_{22}^{-1}C_{21}|}{|C_{11}|} = \frac{|E|}{|E+H|} \tag{9.5.10}$$

where

$$\begin{aligned} E &= C_{11}-C_{12}C_{22}^{-1}C_{21} \\ H &= C_{12}C_{22}^{-1}C_{21} \end{aligned} \tag{9.5.11}$$

By Theorem 23, section 3.6; E and H have independent distributions such that

$$\begin{aligned} E &\sim \mathcal{W}_p(I;\nu-q:(0)) \\ H &\sim \mathcal{W}_p(I;q:(0)) \end{aligned} \tag{9.5.12}$$

so that

$$\begin{cases} \eta|H_o \sim U_{p:q:\nu-q} \\ \eta|H_o \sim U_{q:p:\nu-p} \end{cases} \tag{9.5.13}$$

Exact results for the density and percentage points are given (section 4.6) when p = 1 or 2 or q = 1 or 2. Approximations to the significance level of $U_{m:\nu_1:\nu_2}$ are given in equation 4.6.18.

It is noted that when p=q=1 (m=2), η becomes the standard statistic for testing the presence of zero correlation in the bivariate case.

EXERCISES

Exercises to Chapter 2.

1. If (x,y) has a bivariate distribution with

$$\mathcal{E}(x) = \mu_x; \quad \mathcal{E}(y) = \mu_y$$

$\mathrm{Var}(x) = \sigma_x^2$; $\mathrm{Cov}(x,y) = \rho\sigma_x\sigma_y$; $\mathrm{Var}(y) = \sigma_y^2$

show that $|\rho| \leq 1$.

[Hint: consider $\mathcal{E}(ax+by+c)^2$ for suitable a,b,c].

2. Extending exercise 1; if $(x_i)_1^m$ have a multivariate distribution with means $(\mu_i)_1^m$ and variance-covariance matrix V, show that V is necessarily positive-definite if there exists no linear function relationship between the $(x_i)_1^m$.

3. If

$$f(x_1,x_2) = \frac{1}{(2\pi)\sqrt{1-\rho^2}} \exp -\frac{1}{2}\left\{\left[\frac{x_1-\mu_1}{\sigma_1}\right]^2 - 2\rho\left[\frac{x_1-\mu_1}{\sigma_1}\right]\left[\frac{x_2-\mu_2}{\sigma_2}\right] + \left[\frac{x_2-\mu_2}{\sigma_2}\right]^2\right\}$$

obtain $a_{11}, a_{12}, a_{21}, a_{22}$ such that

$$y_1 = a_{11}x_1 + a_{12}x_2$$
$$y_2 = a_{21}x_1 + a_{22}x_2$$

are uncorrelated.

Obtain (from first principles) the densities:

$f(x_1)$; $f(x_2)$; $f(x_1|x_2)$ and $f(x_2|x_1)$.

Compare these densities with the densities obtained in the text for

$f(\underline{x}_{(1)})$; $f(\underline{x}_{(2)})$; $f(\underline{x}_{(1)} | \underline{x}_{(2)})$ and $f(\underline{x}_{(2)} | \underline{x}_{(1)})$

in this special case of $m=2$.

4. If $\underline{x} = \begin{bmatrix} \underline{x}_{(1)} \\ \cdots \\ \underline{x}_{(2)} \end{bmatrix} \frown N_m(\underline{\mu};V)$, show that a necessary and sufficient condition for every element of $\underline{x}_{(1)}$ to be independent of every element of $\underline{x}_{(2)}$ is $V_{12} = (0)$.

5. (Extension to Theorem 1). If $\underline{x} \frown N_m(\underline{\mu};V)$ and C is $(h\times m)$ of full rank $h<m$, show that

$$\underline{y} = C\underline{x} \frown N_h(C\underline{\mu};CVC')$$

[Hint: Define Q to be $\overline{m-h} \times m$ such that the t^{th} row of Q is orthogonal to every row of C and to the previous $(t-1)$ rows of Q $(t = 1, 2, \ldots, \overline{m-h})$. Use Theorems 1 and 2.]

6. Suppose $\begin{bmatrix} \underline{x}_{(1)} \\ \cdots \\ \underline{x}_{(2)} \end{bmatrix} \frown N_m \left\{ \begin{bmatrix} \underline{\mu}_{(1)} \\ \cdots \\ \underline{\mu}_{(2)} \end{bmatrix} ; \begin{bmatrix} V_{11} & \vdots & V_{12} \\ \cdots & & \cdots \\ V_{21} & \vdots & V_{22} \end{bmatrix} \right\}$

give the density of

$$\begin{bmatrix} \underline{x}_{(1)} \\ \cdots \\ \underline{y} \end{bmatrix} = \begin{bmatrix} I & \vdots & (0) \\ \cdots & & \cdots \\ B & \vdots & C \end{bmatrix} \begin{bmatrix} \underline{x}_{(1)} \\ \cdots \\ \underline{x}_{(2)} \end{bmatrix}$$

and demonstrate that B and C may be so chosen that the dispersion matrix for

$$\begin{bmatrix} \underline{x}_{(1)} \\ \cdots \\ \underline{y} \end{bmatrix} \quad \text{reduces to} \quad \begin{bmatrix} V_{11} & \vdots & (0) \\ \cdots & & \cdots \\ (0) & \vdots & I \end{bmatrix}$$

Hence give the density of $\underline{x}_{(1)}$.

7. With $\begin{bmatrix}\underline{x}_{(1)}\\ \cdots\cdots \\ \underline{x}_{(2)}\end{bmatrix}$ defined as in question 6, give the density of $\begin{bmatrix}\underline{x}_{(1)}\\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \underline{x}_{(2)} - V_{21}V_{11}^{-1}\underline{x}_{(1)}\end{bmatrix}$. Use this and the marginal density of $\underline{x}_{(1)}$ found in question 6 to give the density of $\underline{x}_{(2)}$ conditional on $\underline{x}_{(1)}$.

8. If

$$\begin{bmatrix}\underline{x}_{(1)}\\ \cdots\cdots \\ \underline{x}_{(2)}\\ \cdots\cdots \\ \underline{x}_{(3)}\end{bmatrix} \frown N_m\left\{\begin{bmatrix}\underline{\mu}_{(1)}\\ \underline{\mu}_{(2)}\\ \underline{\mu}_{(3)}\end{bmatrix} ; \begin{bmatrix}V_{11} & \vdots & V_{12} & \vdots & V_{13}\\ \cdots & & \cdots & & \cdots \\ V_{21} & \vdots & V_{22} & \vdots & V_{23}\\ \cdots & & \cdots & & \cdots \\ V_{31} & \vdots & V_{32} & \vdots & V_{33}\end{bmatrix}\right\}$$

show that statement A below does not necessarily imply B and vice versa.

A: $f(\underline{x}_{(1)},\underline{x}_{(2)}) = f(\underline{x}_{(1)})f(\underline{x}_{(2)})$

B: $f(\underline{x}_{(1)},\underline{x}_{(2)}|\underline{x}_{(3)}) = f(\underline{x}_{(1)}|\underline{x}_{(3)})f(\underline{x}_{(2)}|\underline{x}_{(3)})$

9. If $(x_1,x_2,x_3,x_4)'$ has dispersion matrix

$$\begin{bmatrix}2 & 1 & 0 & 1\\ 1 & 2 & 1 & 0\\ 0 & 1 & 2 & 1\\ 1 & 0 & 1 & 2\end{bmatrix}$$

obtain the canonical correlation between $\begin{bmatrix}x_1\\ x_2\end{bmatrix}$ and $\begin{bmatrix}x_3\\ x_4\end{bmatrix}$.

10. If $\underline{x}$ is partitioned into $\begin{bmatrix}\underline{x}_{(1)}\\ \cdots\cdots \\ \underline{x}_{(2)}\end{bmatrix} \frown N_m\left\{\begin{bmatrix}\underline{\mu}_{(1)}\\ \underline{\mu}_{(2)}\end{bmatrix} ; \begin{bmatrix}V_{11} & \vdots & V_{12}\\ \cdots & & \cdots \\ V_{21} & \vdots & V_{22}\end{bmatrix}\right\}$

and x_i is an element of $\underline{x}_{(1)}$, obtain $\underline{a}_0$ such that

Var $(x_i - \underline{a}'\underline{x}_{(2)})$ is a minimum over choice of $\underline{a}$ at $\underline{a} = \underline{a}_0$. Obtain the correlation between x_i and $\underline{a}_0'\underline{x}_{(2)}$ and compare with the coefficient of multiple correlation.

11. With $\underline{x}_{(1)}$, $\underline{x}_{(2)}$ and x_i defined as in question 10 and $R_{i.(2)}$ the multiple correlation coefficient between x_i and $\underline{x}_{(2)}$, show that $R_{i.(2)}$ is invariant with respect to linear transformations on $\underline{x}_{(2)}$ is replaced by $K\underline{x}_{(2)}$ when K is non-singular. (Does K have to be square?)

12. If $m = 3$, show that

$\rho_{12.3} = V_{12.3} \Big/ V_{22.3}^{\frac{1}{2}}$ $\rho_{23.1} = V_{23.1} \Big/ V_{33.1}^{\frac{1}{2}}$ etcetera.

13. If $m = 3$, show that

$$1-R^2_{1.23} = (1-\rho^2_{12})(1-\rho^2_{13.2})$$

where

$$\rho_{12} = V_{12} \Big/ (V_{11}V_{22})^{\frac{1}{2}}$$

$$\rho_{13.2} = V_{13.2} \Big/ (V_{11.2}V_{33.2})^{\frac{1}{2}}.$$

14. If the dispersion matrix V of the vector observation $\underline{x}' = (x_1, x_2, \ldots, x_m)$ is such that $V_{ii} = 1$; $V_{ij} = \rho$ (all i,j; $i \neq j$), show that

$$\rho \geq -1/(m-1)$$

unless there exists some linear relationship amongst the $(x_i)_1^m$.

15. If $\begin{bmatrix} x \\ y \end{bmatrix} \quad N_2 \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} ; \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} \right\}$

obtain Pr $[|x| > |y|]$.

[Hint: Use the result of question 3 and transform to new independent variables.]

16. If $(\underline{x}_j)_1^n$ are a set of vector variables from $N_m(\underline{\mu};V)$ obtain the density of $(\bar{\underline{x}},\underline{u}_1,\underline{u}_2,\ldots,\underline{u}_{n-1})$ defined by

$$\begin{aligned}
\underline{x}_1 &= \bar{\underline{x}} + \frac{\underline{u}_1}{\sqrt{2.1}} + \frac{\underline{u}_2}{\sqrt{3.2}} + \frac{\underline{u}_3}{\sqrt{4.3}} + \frac{\underline{u}_4}{\sqrt{5.4}} + \cdots\cdots + \frac{\underline{u}_{n-1}}{\sqrt{n(n-1)}}\\
\underline{x}_2 &= \bar{\underline{x}} - \frac{\underline{u}_1}{\sqrt{2.1}} + \frac{\underline{u}_2}{\sqrt{3.2}} + \frac{\underline{u}_3}{\sqrt{4.3}} + \frac{\underline{u}_4}{\sqrt{5.4}} + \cdots\cdots + \frac{\underline{u}_{n-1}}{\sqrt{n(n-1)}}\\
\underline{x}_3 &= \bar{\underline{x}} \qquad\qquad - \frac{2\underline{u}_2}{\sqrt{3.2}} + \frac{\underline{u}_3}{\sqrt{4.3}} + \frac{\underline{u}_4}{\sqrt{5.4}} + \cdots\cdots + \frac{\underline{u}_{n-1}}{\sqrt{n(n-1)}}\\
\underline{x}_4 &= \bar{\underline{x}} \qquad\qquad\qquad\qquad - \frac{3\underline{u}_3}{\sqrt{4.3}} + \frac{\underline{u}_4}{\sqrt{5.4}} + \cdots\cdots + \frac{\underline{u}_{n-1}}{\sqrt{n(n-1)}}\\
\\
\underline{x}_n &= \bar{\underline{x}} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad - \frac{(n-1)\underline{u}_{n-1}}{\sqrt{n(n-1)}}
\end{aligned}$$

Hence show that

$$\left.\begin{aligned}\bar{\underline{x}} &\frown N_m(\underline{\mu};\tfrac{1}{n}V)\\ \underline{u}_j &\frown N_m(\underline{0};V)\end{aligned}\right. ; \qquad \left\{\bar{\underline{x}},\ (u_j)_1^{n-1}\right\} \text{ are mutually independent.}$$

17. Suppose $(\underline{x}_{ij})_{i=1}^{k}{}_{j=1}^{n_i}$ are a set of independently distributed vector variables with

$$\underline{x}_{ij} \frown N_m(\underline{\theta}_i : V) \qquad i = 1, 2,\ldots, k;\ \ j = 1, 2,\ldots, n_i,$$

use exercise 16 to show that if $\overline{\underline{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \underline{x}_{ij}$, then

$$\sum_i \sum_j (x_{ij}-\overline{\underline{x}}_i)(\underline{x}_{ij}-\overline{\underline{x}}_i)' \sim \sum_{i=1}^{k} \sum_{j=1}^{n_i-1} \underline{u}_{ij}\underline{u}_{ij}'$$

where $(\underline{u}_{ij})_{i=1}^{k}{}_{j=1}^{n_i-1}$ are independent and

$$\underline{u}_{ij} \frown N_m(\underline{0};V) \qquad i=1, 2,\ldots,k; \quad j=1, 2,\ldots,n_i-1.$$

Show also that is $\overline{\underline{x}}_{..} = (\sum_i \sum_j \underline{x}_{ij}) \Big/ (\sum_i n_i)$, then

$$\sum_{i=1}^{k} (\overline{\underline{x}}_i-\overline{\underline{x}}_{..})(\overline{\underline{x}}_i-\overline{\underline{x}}_{..})n_i \sim \sum_{i=1}^{k-1} \underline{n}_i\underline{n}_i'$$

and give the density of $(\underline{n}_i)_{i=1}^{k-1}$

18. Suppose $(\underline{x}_{ij})_{i=1;\,j=1}^{r;\ c}$ are a set of independently distributed vector variables with

$$\underline{x}_{ij} \frown N_m(\underline{\mu}+\underline{\alpha}_i+\underline{\beta}_j: V) \qquad i = 1, 2,\ldots, r; \quad j = 1, 2,\ldots, c,$$

show that

$$\sum_{i=1}^{r} \sum_{j=1}^{c} (\underline{x}_{ij}-\overline{\underline{x}}_{i.}-\overline{\underline{x}}_{.j}+\overline{\underline{x}}_{..})(\underline{x}_{ij}-\overline{\underline{x}}_{i.}-\overline{\underline{x}}_{.j}+\overline{\underline{x}}_{..})' \sim \sum_{i=1}^{r-1} \sum_{j=1}^{c-1} \underline{n}_{ij}\underline{n}_{ij}'$$

where $(\overline{\underline{x}}_{i.} = \frac{1}{c}\sum_{j=1}^{c}\underline{x}_{ij})_{i=1}^{r}$; $(\overline{\underline{x}}_{.j} = \frac{1}{r}\sum_{i=1}^{r}\underline{x}_{ij})_{j=1}^{c}$

and $\overline{\underline{x}}_{..} = \frac{1}{rc} \sum_{i=1}^{r} \sum_{j=1}^{c} \underline{x}_{ij}$

and $(\underline{n}_{ij})_{i=1}^{r-1}{}_{j=1}^{c-1}$ are a set of mutually independent vector variables such that $\underline{n}_{ij} \frown N_m(\underline{0}; V)$.

[Hint: For fixed i use exercise 16 transforming $(\underline{x}_{ij})_{j=1}^{c}$ to $(\overline{\underline{x}}_{i.}, \underline{u}_{i1}, \underline{u}_{i2}, \ldots, \underline{u}_{i:c-1})$; then perform a similar transformation over a set of r normally distributed vectors.]

19. If $\underline{x} \frown N_m(\underline{\mu}: V)$, show that the quantity $(\underline{x}-\underline{\mu}_o)'V^{-1}(\underline{x}-\underline{\mu}_o)$ has a noncentral χ^2-distirbution and give the degrees of freedom and the noncentrality parameter.
[Hint: Assume the existence of nonsingular M such that $MVM' = I$.]

20. Show how the result of exercise 19 could be used to make a confidence statement on $\underline{\mu}$ when $\underline{x} \frown N_m(\underline{\mu}; V)$ and V is known.

Exercises to Chapter 3.

1. If $(\underline{x}_j)_{j=1}^{n}$ are independently drawn from the population $N(\underline{\mu}; V)$, use Theorem 15 to obtain the distributions of

$$\overline{\underline{x}}\,\overline{\underline{x}}' \quad \text{and} \quad \sum_{j=1}^{n}(\underline{x}_j-\overline{\underline{x}})(\underline{x}_j-\overline{\underline{x}})'$$

where $\overline{\underline{x}} = \frac{1}{n}\sum_{j=1}^{n}\underline{x}_j$

Is $\overline{\underline{x}}$ distributed independently of $\sum_{j=1}^{n}(\underline{x}_j-\overline{\underline{x}})(\underline{x}_j-\overline{\underline{x}})'$? Compare this result with exercise 16 to Chapter 2.

2. Use Theorem 15 to prove the main results of exercises 17 and 18 to Chapter 2.

3. If $X = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{\ })$ has a density of the form

$$\phi(XX')$$

show that with $C = XX'$ the density of the mathematically independent elements of C have the joint density

$$\frac{|C|^{\frac{1}{2}(\nu-m-1)}\phi(C)}{\pi^{\frac{1}{2}m(m-1+2\ )}\prod_{j=1}^{m}\Gamma\left(\frac{\nu+1-j}{2}\right)}$$

4. If $\{(x_{1i}, x_{2i})\}_1^n$ are drawn from the general bivariate density, show that

$$C_{22} - \frac{C_{12}^2}{C_{11}} \quad \text{and} \quad \frac{C_{12}^2}{C_{11}}$$

have independent χ^2-distributions conditional on C_{11} where

$$C_{\alpha\beta} = \sum_1 (x_{\alpha i} - \overline{x}_{\alpha .})(x_{\beta i} - \overline{x}_{\beta .}) \qquad \begin{matrix}\alpha \\ \beta\end{matrix} = 1,2$$

$$\overline{x}_{\alpha} = \frac{1}{n}\sum_1 x_{\alpha i} \qquad \alpha = 1,2$$

Give the density of

$$\frac{r^2}{1-r^2} = \frac{C_{12}^2/C_{11}}{C_{22}-C_{12}^2/C_{11}} \qquad \underline{\text{conditional}} \text{ on } C_{11}$$

and hence obtain the nonconditional distribution of r^2.

5. If $\{\underline{x}_j\}_1^n$ are independently drawn from $N_m(\underline{\mu}; I)$ and

$$r_{\alpha i} = C_{\alpha i} \Big/ \sqrt{C_{\alpha\alpha}C_{ii}} \qquad i = 1,2,\ldots,\alpha-1,\alpha+1,\ldots,m$$

with

$$C_{\alpha\beta} = \left(\sum_j (\underline{x}_j - \overline{\underline{x}})(\underline{x}_j - \overline{\underline{x}})'\right)_{\alpha\beta}$$

show that for fixed a, $\{r_{ai}\}_{i\neq a}$ are independently distributed.

[Hint: As a first step consider the distributions with the a^{th} elements of $\{\underline{x}_j\}$ held fixed.]

6. If $\{\underline{x}_j\}_1^n$ are independently drawn from $N_m(\underline{\mu}; V)$, show that $\mathcal{E}(C) = (n-1)V$ when

$$C = \sum_{j=1}^{n} (x_j-\overline{\underline{x}})(\underline{x}_j-\overline{\underline{x}})'.$$

7. Obtain the Wishart density for the case m=2. If $\left\{\underline{u}_j = \begin{bmatrix} u_{1j} \\ u_{2j} \end{bmatrix}\right\}_{j=1}^{n}$ are independently drawn from $N_2(0; V)$, write down the joint density of

$$C_{11} = \sum_j (u_{1j}-\overline{u}_1)^2$$

and

$$C_{12} = \sum_j (u_{1j}-\overline{u}_1)(u_{2j}-\overline{u}_2)$$

conditonal on $\{\underline{u}_{2j}\}_1^n$. Hence obtain the joint density of C_{11}, C_{12}, and C_{22} when $C_{22} = \sum_j (u_{2j}-\overline{u}_2)^2$.

8. Since $\frac{1}{n-1}C$ is unbiassed for V where

$$C = \sum_j (\underline{x}_j-\overline{\underline{x}})(\underline{x}_j-\overline{\underline{x}})'$$

when $\{\underline{x}_j\}_1^n$ are independently drawn from $N_m(\underline{\mu}; V)$, it would seem not unreasonable to estimate functions of V by corresponding functions of $\frac{1}{n-1}C$. How then would one estimate $R^{2\,*}_{(1)(2)}$? If

$\hat{R}^2_{(1)(2)}$ is the estimator, show that $\hat{R}^2_{(1)(2)}$ is the maximum value of R^2 for which

$$|C_{11}-C_{12}C_{22}^{-1}C_{21} - \frac{R^2}{1-R^2}\, C_{12}C_{22}^{-1}C_{21}| = 0.$$

What can be said about the densities of $C_{11}-C_{12}C_{22}^{-1}C_{21}$ and $C_{12}C_{22}^{-1}C_{21}$ when V partitions as

$$V = \left[\begin{array}{c|c} V_{11} & (0) \\ \hline (0) & V_{22} \end{array}\right]$$

[* let $\underline{x}_j$ partition into $\left[\begin{array}{c} \underline{x}_{(1)j} \\ \hline \underline{x}_{(2)j} \end{array}\right]$; C and V are partitioned in accordance.]

9. If $\left\{\begin{bmatrix} x_{1j} \\ x_{2j} \end{bmatrix}\right\}_{j=1}^{n}$ are a set of n bivariate observations, give geometric meaning to

$$\sum_{j=1}^{n} x_{1j}^2,\ \sum x_{2j}^2, \text{ and } \sum x_{1j}x_{2j} \Big/ \sqrt{\sum x_{1j}^2 \sum x_{2j}^2}$$

[Hint: Regard the observations as two points in n dimensional space rather than n points in two dimensional space.]

10. If A is symmetric, idempotent, nxn of rank r, show that the equation

$$\underline{z}'A\underline{z} = c^2$$

is a sphere of dimension r, radius c in the n-space of $\underline{z}$.

11. Give the univariate equivalents of the theorems of Chapter 3 whenever this is possible.

12. Give the bivariate statements (m=2) of the theorems of

Chapter 3 whenever possible.

13. $$\left.\begin{array}{l} \text{If } X \sim N_{m\times n}((0);\, V \otimes I) \\ A \text{ is idempotent of rank } r \\ \{\underline{a}_t\}_1^k \text{ are a set of vectors} \end{array}\right\}$$

state the conditions under which

$$XAX' \text{ and } \{X\underline{a}_t\}_{t=1}^k$$

have mutually independent distribution. What is the maximum value k can take?

Exercises to Chapter 4.

1. Obtain the h^th^ moment of (central) $U_{m:v_1:v_2}$ and of $U_{v_1:m:v_1+v_2-m}$. Show by direct comparison that

$$\mathcal{E}(U^h_{m:v_1:v_2}) = \mathcal{E}(U^h_{v_1:m:v_1+v_2-m}) .$$

2. By identifying moments, show that

$$U_{2r:v_1:v_2} \sim \prod_{j=1}^{r} Y_j^2$$

where $\{Y_j\}_1^r$ are mutually independent variates with densities:

$$p(Y_j) = Y_j^{v_2-2j}(1-Y_j)^{v_1-1} \Big/ B(v_2-2j+1;\, v_1).$$

(See Theorem 36).

3. By identifying moments, show that

$$U_{2r+1:\nu_1:\nu_2} \sim \prod_{j=1}^{r} Y_j^2 \; X$$

with Y_j defined as in question 2 and X independent of $\{Y_j\}$ with density

$$p(X) = X^{\frac{1}{2}(\nu_2+1-m)-1}(1-X)^{\frac{1}{2}\nu_1-1} \Big/ B(\tfrac{1}{2}(\nu_2+1-m);\ \tfrac{1}{2}\nu_1).$$

(See Theorem 37).

4. If y is a random variable distributed on $[0,1]$ with density

$$p(y) = y^{\alpha-1}(1-y)^{\beta-1} \Big/ B(\alpha,\beta)$$

obtain an approximation to the percentage points of y in terms of the percentage points of χ^2-variates.
[Use the techniques described in section 4.6.]

5. Show that, for fixed c and a at choice

$$I(a,c) = \int_a^{a+c} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du$$

is maximized when $a = -\frac{1}{2}c$ [Do this by investigating the zero of $\frac{d}{da} I(a,c)$]. Compare this result with Lemma 3 (section 4.7).

6. Suppose that $U_{(i)}$ has the distribution $U_{m_i:\alpha_i:\beta_i}$ $(i=1,2)$ and that $U_{(1)}$ and $U_{(2)}$ are independent. Is it possible that $U = U_{(1)}U_{(2)}$ is distributed as $U_{m:\alpha:\beta}$? If so, give the relationships between the 9 parameters.

Exercises to Chapter 5.

1. Obtain the characteristic function of a (univariate) normal deviate with mean μ and variance σ^2.

2. Use the result of question 1 to obtain the joint characteristic function of a vector of random variables. [Assume without justification that linear transformations on complex arguments are valid.]

3. Obtain the joint characteristic function of $X = XX'$ when $X \frown N_{m\times n}(\mu; V \otimes I)$ assuming the characteristic function of u^2 when $u \frown N_1(\xi, \sigma^2)$.

4. A vector of random variables $\underline{x}$ is such that $\underline{a}'\underline{x}$ has a non-degenerate normal density for any $\underline{a}$ $(\underline{a} \neq \underline{0})$; use the theory of characteristic functions to show that the density of $\underline{x}$ must be of the form

$$f(x_1, x_2, \ldots, x_m) = \frac{|V|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}m}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'V^{-1}(\underline{x}-\underline{\mu})}.$$

5. Discuss the class of linear combinations of elements of XX' which have χ^2-distributions when $X \frown N_{m\times n}(\mu; V \otimes I)$.

6. Use the method of characteristic function to prove that the sum of two independent Wishart variates (with a common dispersion matrix) is a Wishart variate.

7. If $u_j \frown N_1(\mu_j; 1)$ obtain the characteristic function for

$\sum_1^v u_j^2$ when the $(u_j)_1^v$ are independent. Hence or otherwise express $\Pr(\sum_1^v u_j^2 \leq t)$ as a weighted infinite sum of incomplete gamma integrals.

8. If $W \sim \mathcal{W}_m(V;v:A)$ and A is of rank one, show that the density of the functionally independent elements of W is a function only of $|V^{-1}W|$, $\mathrm{tr}(V^{-1}W)$ and $\mathrm{tr}(AW)$.
[Note: T.W. Anderson (1946) Annals of Mathematical Statistics pp 409-431.]

9. Show by the use of characteristic functions that if $(\underline{x}_j)_1^n$ are mutually independent such that

$$\mathcal{E}(\underline{x}_j) = \underline{\mu} \qquad j = 1, 2, \ldots, n$$

$$\mathcal{E}(\underline{x}_j - \underline{\mu})(\underline{x}_j - \underline{\mu})' = V \qquad j = 1, 2, \ldots, n$$

and if further $\underline{\bar{x}} = \frac{1}{n}\sum_1^n x_j$ and $C = \sum_{j=1}^n (\underline{x}_j - \underline{\bar{x}})(\underline{x}_j - \underline{\bar{x}})'$ have independent distributions, then $\underline{x}_j$ must have a multivariate normal density.
[Hint: Consider the univariate situation first.]

Exercises to Chapter 6.

1. Obtain by direct methods the Jacobian of equation (6.2.13) for the case m=2.

2. The Haar Invariant density on the m^2 variates $(a_{ij})_{1;1}^{m;m}$ can be written

$$f((\ell_{ij})_{1,1}^{m,m}) = \frac{|L'L|^{\frac{1}{2}(v-m)} e^{-\frac{1}{2}\mathrm{tr}(L'L)}}{2^{\frac{1}{2}mv} \pi^{\frac{1}{2}m^2}} \prod_{j=1}^{m} \frac{\Gamma(\frac{1}{2}(m+1-j))}{\Gamma(\frac{1}{2}(v+1-j))}$$

over the region $-\infty<\ell_{ij}<\infty)_{i=1;\, j=1}^{m;\ m}$ where $L = (\ell_{ij})$

Obtain the density of $L^* = PL$ when P is orthogonal. Show that the above is indeed a density function.

3. How is L related to a matrix of normal deviates when L has the Haar Invariant density?

4. If L has the Haar Invariant density, obtain the density of L'L.

5. If $W_{(i)} \frown \mathcal{W}_2(V; v_i : (0))$ (N.B. m=2) and $W_{(1)}$, $W_{(2)}$ are independent and λ_1, λ_2 are the two roots of

$$|W_{(1)} - \lambda W_{(2)}| = 0, \quad (\lambda_1 > \lambda_2)$$

obtain the densities of

(a) λ_1; (b) λ_2; (c) $\lambda_1 + \lambda_2$.

6. If $W \frown \mathcal{W}_2(V; v : (0))$ (N.B. m=2), obtain the joint density on the characteristic roots of W.

7. If λ_1, λ_2 $(\lambda_1 > \lambda_2)$ are the roots in problem 6, obtain the density of $\sqrt{\lambda_1 \lambda_2} \Big/ \frac{1}{2}(\lambda_1 + \lambda_2)$

(a) when $V = \sigma^2 I$

(b) for general V (V is 2x2).

8. The "sphericity criterion" is the ratio of geometric mean to arithmetic mean of the latent roots of W when $W \frown \mathcal{W}_m(V; v; (0))$ the "sphericity" concept implying $V = \sigma^2 I$. Obtain moments of this criterion.

Exercises to Chapter 7.

1. A sequence of n_i vector observations are made on the i^{th} of g populations. If it be assumed that

$$\underline{x}_{ij} \frown N_m(\underline{\mu}_i; V) \qquad j = 1, 2, \ldots, n_i; \quad i = 1, 2, \ldots, g$$

and $(\underline{x}_{ij})_{1\ 1}^{g\ n_i}$ are mutually independent, obtain the likelihood ratio tests of the following hypothesis when V is unknown:

(a) $\underline{\mu}_1 = \underline{\mu}_2 = \cdots = \underline{\mu}_{(g)}$ (unknown)

(b) $\beta_1\underline{\mu}_1 + \beta_2\underline{\mu}_2 + \cdots + \beta_g\underline{\mu}_{(g)} = \underline{\xi}$ ($\underline{\xi}$ and $(\beta_i)_1^g$ specified).

(c) $\{M\underline{\mu}_i = \underline{0}\ (s \times 1)\}_{i=1}^{g}$ when M is specified.

Give the distribution type in each case under H_0.

2. How would your test functions and their null distributions differ (if at all) if in question 1 it were known that V is a diagonal matrix with unknown diagonal elements?

3. A vector of measurements $\underline{x}_{ij}$ is made on a system at time t_i ($j = 1, 2, \ldots, n_i$; $i = 1, 2, \ldots, g$). It is required to test that no element in $\mathcal{E}(\underline{x}) = \underline{\mu}(t)$ is a polynomial of higher order in than p. Obtain the likelihood ratio test function if the

$\underline{x}_{ij}$ have a common dispersion matrix. [$\underline{\mu}(t)$ is to be considered as a vector of different polynomials in t with unknown coefficients.]

4. In a two-way cross-classification with equal replication of vector observations, it is required to test that the "row-effect" and "column-effect" are additive. Obtain the likelihood ratio test function. Give the distribution type of a suitable function of the test function.

5. Why may any ANOVA in the univariate case be generalized to the multivariate case by simply replacing a square (univariate) by the product of a column vector by its transpose (multivariate)?

6. If ${}_{(i)}X \frown N_{m\times n_i}({}_{(i)}B\,{}_{(i)}\Xi;\ {}_{(i)}V)$ $(i=1,2)$ obtain a test of the hypothesis that the ${}_{(i)}X$ were drawn from the same population (i.e., ${}_{(1)}B = {}_{(2)}B;\ {}_{(1)}V = {}_{(2)}V$).

7. Two populations are such that

$$\underline{x} \frown N_m({}_{(1)}\underline{\mu};\ {}_{(1)}V) \text{ if } \underline{x} \text{ is from population 1}$$

$$\underline{x} \frown N_m({}_{(2)}\underline{\mu};\ {}_{(2)}V) \text{ if } \underline{x} \text{ is from population 2.}$$

If ${}_{(1)}\underline{\mu};\ {}_{(2)}\underline{\mu};\ {}_{(1)}V$ and ${}_{(2)}V$ are known, obtain the "best" vector $\underline{\ell}^{*}$ to be used in making a decision on whether $\underline{x}$ belongs to population 1 or 2, on the basis of the scalar quantity $\underline{\ell}'\underline{x}$.
[* "best" may be defined in a variety of ways; use the vector which minimizes the probability of misclassification.]

8. In problem 7 use a likelihood ratio approach and compare the results.

9. Given $X \frown N_{m \times n}(B\xi;\ V \times I)$ with V known, it is required to test $BA = \Gamma$ with A and Γ specified. It is suggested that for suitable $\underline{a}$ the transformation

$$\underline{y}' = \underline{a}'X$$

is made so that $\underline{y}' \frown N_{1 \times n}((\underline{a}'B\xi);(\underline{a}'V\underline{a})I)$. Suggest how $\underline{y}'$ could be used to test H_0: $BA = \Gamma$. What is the distribution of your test function? How should $\underline{\alpha}$ be chosen? [These considerations are at the basis of the modern union-intersection proceedures.]

10. A measure of height is made on the oldest m male sibs in each of n families. It is required to test the hypothesis that there is no difference in mean height of these sibs. Give full details for the execution of such a test.

Exercises to Chapter 9.

1. It is thought that the dispersion matrix for a given m-variate population (V, say) is of the form

$$(V)_{ij} = \alpha\delta_{ij} + \beta$$

where α and β are unknown constants and δ_{ij} is the Kronecker delta. The mean vector ($\underline{\mu}$, say) is entirely unknown. Suggest ways of testing this hypothesis.

2. Extend section 9.5 by considering the likelihood ratio test of the hypothesis that the dispersion matrix is of the form

$$V = \begin{bmatrix} V_{11} & (0) & (0) & \cdots\cdots & (0) \\ (0) & V_{22} & (0) & \cdots\cdots & (0) \\ (0) & (0) & V_{33} & \cdots\cdots & (0) \\ \vdots & \vdots & \vdots & & \\ (0) & (0) & (0) & \cdots\cdots & V_{kk} \end{bmatrix}$$

where the dimension of V_{ii} is, say, $p_i \times p_i$ $(\sum_i p_i = m)$. What are the moments of the test statistic (or a convenient monotone function of it), if the hypothesis is true?

3. Observational vectors $\underline{x}$ can be made independently on a population. It is the hypothesis that

$$\begin{cases} x_1 = z_1 + w_1 \\ x_2 = x_1 + z_2 + w_2 \\ x_3 = x_2 + z_3 + w_3 \\ \vdots \\ x_m = x_{m-1} + z_m + w_m \end{cases}$$

where $(w_i)_1^m$ are unknown constants and the $(z_i)_1^m$ are independent normal variates with zero mean and the same variance. Suggest possible tests of this hypothesis.

4. Obtain the likelihood ratio test for section 9.4 in the case $m=1$. (Test of homoscedasticity in the univariate case.) Use the definition of u equation (9.4.16) in the case $m=1$ and obtain the moment generating function of $-2\rho \log u$. Choose a suitable value of ρ (ρ_o, say) and give an approximation to the distribution $-2\rho_o \log u$.

5. It is thought that a vector of observations $\underline{x}$ is related to $\underline{u} \frown N_p(\underline{\mu};\ V)$ and $\underline{z} \frown N_m(\underline{0};\ \sigma^2 I)$ in the manner:

$$\underline{x} = A\underline{u} + \underline{z}$$

If $\underline{u}$ and $\underline{z}$ have independent densities and A is of rank $p(<m)$, show that the dispersion matrix of $\underline{x}$ has a latent root repeated at least $(m-p)$ times. Suggest a possible test of the structure in the case $p=1$. [These considerations are at the basis of factor analysis.]

APPENDIX A

SOME RESULTS IN THE THEORY OF MATRICES AND DETERMINANTS

A.1 Introduction

It will be assumed that the reader has a working knowledge of the algebra of matrices. Many theorems are repeatedly required in the development of multivariate theory; for example

(i) given a symmetric matrix V, there exist an orthogonal matrix P such that $PVP' = D$ with D a diagonal matrix such that $\{d_{ii}\}$ accounts for all the roots of the determinantal equation $|V-dI| = 0$.

(ii) if V is symmetric and positive definite, there exists (non-singular) M such that $MVM' = I$.

The proof of these theorems are standard to any good text book in the matrix algebra and no proofs will be given here.

Other results which we will need are, however, not often to be found discussed (or even quoted) in standard college text books. Some reults needed in Chapter 1 of these notes are given in this appendix and since a study of the proofs of these results are likely to prove helpful in refamiliarizing the reader with the algebra of matrices; these proofs are offered also.

A.2 The clockwise rule.

Let V be the matrix (v_{ij}) (mxm) and partition it:-

$$V = \begin{bmatrix} V_{11} & \vdots & V_{12} \\ \cdots & & \cdots \\ V_{21} & \vdots & V_{22} \end{bmatrix} \qquad \text{(A.2.1)}$$

with V_{11}, pxp and V_{22}, qxq; so that V_{12} is pxq and V_{21} is qxp, (p+q = n).

Consider a matrix B of the form

$$\text{(A.2.2)} \qquad B = \left[\begin{array}{c:c} I & (0) \\ \hdashline M & I \end{array}\right]$$

the partitioning being as that for V. Then

$$\text{(A.2.3)} \qquad BVB' = \left[\begin{array}{c:c} V_{11} & V_{11}M'+V_{12} \\ \hdashline MV_{11}+V_{21} & MV_{11}M'+V_{21}M'+MV_{12}+V_{22} \end{array}\right]$$

if $|V_{11}| \neq 0$ we choose now to define M by

$$\text{(A.2.4)} \qquad M = -V_{21}V_{11}^{-1}$$

and obtain

$$\text{(A.2.5)} \qquad BVB' = \left[\begin{array}{c:c} V_{11} & V_{12}-V_{11}V_{11}'^{-1}V_{21}' \\ \hdashline (0) & V_{22}-V_{21}V_{11}^{-1}V_{12} \end{array}\right]$$

Since $|B| = 1$, $|BVB'| = |V|$. Equating this to the determinant of the right-hand side of (A.2.5) we get

$$\text{(A.2.6)} \qquad |V| = |V_{11}||V_{22}-V_{21}V_{11}^{-1}V_{12}|. \quad (\text{if } |V_{11}| \neq 0).$$

A slightly different approach (i.e. writing an upper triangle matrix B) gives

$$\text{(A.2.7)} \qquad |V| = |V_{22}||V_{11}-V_{12}V_{22}^{-1}V_{21}| \quad (\text{if } |V_{22}| \neq 0).$$

The rule implied by equations (A.2.6) and (A.2.7) is sometimes referred to as "the clockwise rule". Incidentally, this value gives quite a quick way of evaluating (3x3) or (4x4) determinants.

Although the rule applies to any square matrix V, we shall apply it in the case V symmetric throughout Chapter 1. Also, when V is positive definite, the non-singularity of V_{11} is of course ensured for any partitioning $p<n$.

A.3 The inverse of a partitioned matrix.

Let V be partitioned as in section A.2 and suppose V, V_{11}, and V_{22} are non-singular. If

$$A = \begin{bmatrix} A_{11} & \vdots & A_{12} \\ \cdots & & \cdots \\ A_{21} & \vdots & A_{22} \end{bmatrix} = V^{-1} \tag{A.3.1}$$

the partitioning of A being as that for V then since

$$\begin{bmatrix} V_{11} & \vdots & V_{12} \\ \cdots & & \cdots \\ V_{21} & \vdots & V_{22} \end{bmatrix}\begin{bmatrix} A_{11} & \vdots & A_{12} \\ \cdots & & \cdots \\ A_{21} & \vdots & A_{22} \end{bmatrix} = \begin{bmatrix} I & \vdots & (0) \\ \cdots & & \cdots \\ (0) & \vdots & I \end{bmatrix} \tag{A.3.2}$$

we have

$$\begin{aligned} V_{11}A_{11}+V_{12}A_{21} &= I & \text{i.e.}\quad A_{11} &= V_{11}^{-1}-V_{11}^{-1}V_{12}A_{21} & (|V_{11}|\neq 0) \\ V_{11}A_{12}+V_{12}A_{22} &= (0) & \text{i.e.}\quad A_{12} &= -V_{11}^{-1}V_{12}A_{22} & (|V_{11}|\neq 0) \\ V_{21}A_{11}+V_{22}A_{21} &= (0) & \text{i.e.}\quad A_{21} &= -V_{22}^{-1}V_{21}A_{11} & (|V_{22}|\neq 0) \\ V_{21}A_{12}+V_{22}A_{22} &= I & \text{i.e.}\quad A_{22} &= V_{22}^{-1}-V_{22}^{-1}V_{21}A_{12} & (|V_{22}|\neq 0). \end{aligned} \tag{A.3.3}$$

From the first and third of these equations

$$(V_{11}-V_{12}V_{22}^{-1}V_{21})^{-1} = A_{11} \tag{A.3.4}$$

and from the second and fourth of these equations

(A.3.5) $$(V_{22}-V_{21}V_{11}^{-1}V_{12})^{-1} = A_{22}$$

so that

(A.3.6) $$A_{12} = -V_{11}^{-1}V_{12}(V_{22}-V_{21}V_{11}^{-1}V_{12})^{-1}$$

and

(A.3.7) $$A_{21} = -V_{22}^{-1}V_{21}(V_{11}-V_{12}V_{22}^{-1}V_{21})^{-1}$$

We note that A_{ij} and V_{ij} are interchangeable thus it is equally true (c.f. (a.3.4)) that

$$(A_{11}-A_{12}A_{22}^{-1}A_{21})^{-1} = V_{11}$$

We note also that, by the clockwise rule and using equation (A.3.5)

(A.3.8) $$|V||A_{22}| = |V_{11}|$$

similarly

(A.3.9) $$|V||A_{11}| = |V_{22}|.$$

A.4 The vector and matrix differential operator.

If $K(x) = (k_{ij}(x))$ is a matrix whose elements are functions of x, then we use $\frac{d}{dx} K(x)$ to denote $\frac{d}{dx} k_{ij}(x)$

If $L(x) = (\ell_{ij}(x))$ is a non-singular, mxm matrix whose elements are functions of x then it is easy to show that

$$\frac{d}{dx}|L(x)| = \sum_{j=1}^{m} |\mathcal{D}_j L(x)|$$

where $\mathcal{D}_j$ is an operator which differentiates the j^{th} row of $L(x)$, leaving all other rows the same.

We now define differentiation with respect to a matrix. Let $F\left\{\begin{matrix}\alpha_{11}\cdots\alpha_{1c}\\ \vdots \quad\quad \vdots\\ \alpha_{ri}\cdots\alpha_{rc}\end{matrix}\right\}$ be a scalor, vector, or matrix function of the array of α_{ij} and let $A = (\alpha_{ij})$ be the rxc matrix of the array. We define

$$\frac{dF}{dA} = \frac{dF}{d\alpha_{ij}} \tag{A.4.1}$$

when the α_{ij} are functionally independent; thus if F is a matrix and A is a column vector then dF/dA is a column vector whose elements are matrices.

Some consequences of the definition are

$$\frac{d}{d\underline{\alpha}}(\underline{\alpha}'M) = M \tag{A.4.2}$$

$$\frac{d}{d\underline{\alpha}}(\underline{\alpha}'M\underline{\alpha}) = (M+M')\alpha \tag{A.4.3}$$

when the elements of $\underline{\alpha}$ are functionally independent.

$$\frac{d}{dA}\,\mathrm{tr}[AC] = C' \tag{A.4.4}$$

$$\frac{d}{dA}\,\mathrm{tr}[AA'] = 2A \tag{A.4.5}$$

when the elements of A are functionally independent.

$$\frac{d}{dA}\log|A| = (A')^{-1} \tag{A.4.6}$$

when A is square, non-singular and when the elements of A are functionally independent.

An important result is

Theorem A.1

If C is a fixed, positive definite matrix and A is held to be symmetric but otherwise the elements of A are functionally independent then

$$\max_{A} |A|^{\frac{1}{2}n} e^{-\frac{1}{2}\mathrm{tr}[AC]} = |nC^{-1}|^{\frac{1}{2}n} e^{-\frac{1}{2}nm}$$

the maximum occurring when $A = nC^{-1}$

The stationary values of $\ell(A)$ occur when $\frac{d}{d\alpha_{ij}} \ell(A) = 0 \quad (1 \leq i \leq j \leq m)$.

Since $\log \ell(A)$ is a monotonic increasing function of $\ell(A)$, we will look (for convenience) at the stationary values of $\log \ell(A)$.

$$\frac{d}{d\alpha_{ij}} \log \ell(A) = \frac{n}{|A|}(\text{cof } \alpha_{ij}) - C_{ij} \qquad i \neq j$$

$$= \frac{n}{2|A|}(\text{cof } \alpha_{ii}) - \frac{1}{2}C_{ii} \qquad i = j$$

Equating the differentials to zero, we have

$$\frac{n}{|A|}(\text{cof } \alpha_{ij}) = C_{ij} \qquad \text{all } (i,j)$$

so that

$$nA^{-1} = C.$$

Since there is only one stationary value and since $\ell(A)$ can take the value zero, the stationary value (which is positive) must be a maximum.

This last theorem is of great importance in the solution for the maximum likelihood estimate of the dispersion matrix given a sample of n vectors from a multivariate normal population.

A.5 The Kronecker product of matrices.

Some comments on the Kronecker product of matrices may be helpful since this product is not widely met. If A is (mxm) and B is (nxn) we write

(A.5.1)
$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \ldots & a_{1m}B \\ a_{21}B & a_{22}B & \ldots & a_{2m}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \ldots & a_{mn}B \end{bmatrix}$$

which is an mn x mn matrix.

We see that in general

(A.5.2)
$$A \otimes B \neq B \otimes A$$

We show:

Theorem A.2

If A is (mxm) and B is (nxn) then

$$|A \otimes B| = |A|^n \, |B|^m.$$

Post multiplying $A \otimes B$ by $I \otimes B^{-1}$ and since $|I_m \otimes B^{-1}|$ is $|B|^{-m}$ we have

$$|A \otimes B| = |B|^m |A \otimes I_n|$$

(with I_t the unit matrix of dimension (txt)).

Now if ϕ is a latent root of A, then it is easy to see that $A \otimes I_n$ has a latent root ϕ repeated n times so that

$$|A \otimes I_n| = |A|^n$$

which completes the proof.

Theorem A.2 has an application in the theory of transformation. Suppose X and Y are (mxm) and it is required to transform from the m^2 variables in X to the variables in Y when

$$X = AYB$$

then the Jacobian of this transformation

$$\left\| \begin{array}{cccc:c:cccc} \frac{dx_{11}}{dy_{11}} & \frac{dx_{12}}{dy_{11}} & \cdots & \frac{dx_{1m}}{dy_{11}} & \cdots\cdots & \frac{dx_{m1}}{dy_{11}} & \frac{dx_{m2}}{dy_{11}} & \cdots & \frac{dx_{mm}}{dy_{11}} \\ \frac{dx_{11}}{dy_{12}} & \frac{dx_{12}}{dy_{12}} & \cdots & \frac{dx_{1m}}{dy_{12}} & \cdots\cdots & \frac{dx_{m1}}{dy_{12}} & \frac{dx_{m2}}{dy_{12}} & \cdots & \frac{dx_{mm}}{dy_{12}} \\ & & & & & & & & \\ \frac{dx_{11}}{dy_{mm}} & \frac{dx_{12}}{dy_{mm}} & \cdots & \frac{dx_{1m}}{dy_{mm}} & \cdots\cdots & \frac{dx_{m1}}{dy_{mm}} & \frac{dx_{m2}}{dy_{mm}} & \cdots & \frac{dx_{mm}}{dy_{mm}} \end{array} \right\|$$

is $|A \otimes B| = |A|^m |B|^m$.

APPENDIX B

THE JACOBIAN OF A TRANSFORMATION USED IN CHAPTER 6

B.1 Introduction

In Chapter 6; section 2, we were concerned with the transformation (equation 6.2.11)

$$(B.1.1) \qquad \begin{cases} H = K'D_\phi K \\ E = K'(I-D_\phi)K \end{cases}$$

where all matrices are m×m; D_ϕ is diagonal; H and E are symmetric. The transformation is therefore from

$$\left\{ \{h_{\alpha\beta}\}_{1\leq\alpha\leq\beta\leq m} ; \{e_{\alpha\beta}\}_{1\leq\alpha\leq\beta\leq m} \right\}$$

to

$$\left\{ \{k_{ij}\}_{i=1}^{m}\,{}_{j=1}^{m} ; \{\phi_i\}_{i=1}^{m} \right\}$$

the resulting Jacobian being a determinant of dimension m(m+1). The value of this Jacobian was given (equation (6.2.13)) as

$$2^m|K|^{m+2} \prod_{1\leq i<j\leq m}\prod (\phi_i-\phi_j)$$

and we seek now to prove this value. Straight forward evaluation of the determinant is cumbersome (even when one knows the answer). We report an approach due to Hsu. (Deemar and Olkin, 1951).

B.2 The evaluation.

It is noted that

$$(B.2.1) \qquad \frac{\partial(H,E)}{\partial(K,D_\phi)} = \frac{\partial(H,H+E)}{\partial(K,D_\phi)}$$

Writing

(B.2.2) $$G = H + E$$

we seek the Jacobian of the transformation

(B.2.3) $$\begin{cases} H = K'D_\phi K \\ G = K'K \end{cases}$$

It is noted that the right-hand side of equations (B.2.3) are quadratic in elements of K and linear in elements of D . Suppose a change δK in K and δD_ϕ in D_ϕ takes H into $H+\delta H$ and G into $G+\delta G$ then retaining only terms of first order in the change,

(B.2.4) $$\begin{cases} \delta H = (\delta K)'D_\phi K + K'(\delta D_\phi)K + K'D_\phi(\delta K) \\ \delta G = (\delta K)'K + K'\delta K \end{cases}$$

The derivative of H with respect to an element of K is the coefficient of the corresponding element in δK and similarly the derivative of H with respect to an element of D_ϕ is the coefficient of the corresponding element in δD_ϕ. Thus,

(B.2.5) $$\frac{\partial(H,G)}{\partial(K,D_\phi)} = \frac{\partial(\delta H,\delta G)}{\partial(\delta K,\delta D_\phi)}$$

where the relation between $(\delta H,\delta G)$ and $(\delta K,\delta D_\phi)$ is the linear transformation expressed by equation (B.2.4). It is to be remembered of course that δH and δG are symmetric so that only $2 \times \frac{1}{2}m(m+1)$ equations are implied by (B.2.4). Premultiplying equations (B.2.4) by $(K')^{-1}$ and post multiplying by K^{-1}

(B.2.6)
$$\begin{cases} (K')^{-1}\delta H(K)^{-1} = (K')^{-1}(\delta K')D_\phi + \delta D_\phi + D_\phi(\delta K)K^{-1} \\ (K')^{-1}\delta G(K)^{-1} = (K')^{-1}(\delta K)' + (\delta K)K^{-1} \end{cases}$$

We now consider the transformation

$$(\delta H, \delta G) \rightarrow (\delta K, \delta D_\phi)$$

in three stages

(i) $\left(\delta H; \delta G\right) \rightarrow \left((K')^{-1}\delta H(K)^{-1}; (K')^{-1}\delta G(K)^{-1}\right)$

(ii) $\left((K')^{-1}\delta H(K)^{-1}; (K')^{-1}\delta G(K)^{-1}\right) \rightarrow \left((\delta K)K^{-1}; \delta D_\phi\right)$

(iii) $\left((\delta K)K^{-1}; \delta D_\phi\right) \rightarrow \left(\delta K; \delta D_\phi\right)$.

The Jacobian of the first stage is $|K|^{2(m+1)}$ and of the third stage is $|K^{-1}|^m = |K|^{-m}$, see section 5, appendix A. The Jacobian of the second stage is $|\partial(A,B)/\partial(C\ M)|$ where (compare with (3.2.6))

(B.2.7)
$$\begin{cases} A = C'D_\phi + M + D_\phi C \\ B = C' + C \end{cases}$$

and $M = \delta D_\phi$ is diagonal. Clearly

(B.2.8)
$$\begin{cases} a_{ij} = \phi_j c_{ji} + \phi_i c_{ij} & i<j \\ a_{ii} = 2\phi_i c_{ii} + m_{ii} & \\ b_{ij} = c_{ji} + c_{ij} & i<j \\ b_{ii} = 2c_{ii} & \end{cases}$$

so that

(B.2.9)
$$\frac{\partial[\{a_{ii}\}^m, \{b_{ii}\}^m, \{a_{ij}\}_{i<j}\ \{b_{ij}\}_{i<j}]}{\partial[\{m_{ii}\}^m, \{c_{ii}\}^m, \{c_{ij}\}_{i<j}\ \{c_{ij}\}_{i>j}]}$$

$$= \begin{bmatrix} I & 2D_\phi & (0) & (0) \\ (0) & 2I & (0) & (0) \\ (0) & (0) & L & N \\ (0) & (0) & I & I \end{bmatrix}$$

$$= 2^m |L-N|$$

where

$$\text{(B.2.10)} \qquad L = \frac{\partial(\{a_{ij}\}_{i<j})}{\partial(\{c_{ij}\}_{i<j})}$$

$$= \begin{bmatrix} \phi_1 I_m & (0) & (0) & (0) & .\ . & 0 \\ (0) & \phi_2 I_{m-1} & (0) & (0) & .\ . & 0 \\ (0) & (0) & \phi_3 I_{m-2} & (0) & .\ . & 0 \\ (0) & (0) & (0) & \phi_4 I_{m-3} & .\ . & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & .\ . & \phi_{m-1} I_1 \end{bmatrix}$$

with I_t the unit matrix of dimension t and

$$\text{(B.2.11)} \qquad N = \frac{\partial(\{a_{ij}\}_{i<j})}{\partial(\{c_{ij}\}_{i>j})}$$

$$= \begin{bmatrix} D_\phi^{(2)} & (0) & (0) & \dots & \underline{0} \\ (0) & D_\phi^{(3)} & (0) & \dots & \underline{0} \\ (0) & (0) & D_\phi^{(4)} & \dots & \underline{0} \\ \vdots & \vdots & \vdots & & \vdots \\ \underline{0} & \underline{0} & \underline{0} & \dots & D_\phi^{(m)} \end{bmatrix}$$

with $D^{(t)}$ the diagonal matrix

$$\begin{bmatrix} \phi_t & 0 & 0 & \cdots & 0 \\ 0 & \phi_{t+1} & 0 & \cdots & 0 \\ 0 & 0 & \phi_{t+2} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \phi_m \end{bmatrix}$$

Finally

(B.2.12) $$|L-N| = \prod_{i<j}\prod (\phi_i-\phi_j)$$

The Jacobian of $(H;E) \rightarrow (K;D_\phi)$ is the product of the Jacobians of the three stages and is this

$$|K|^{-m}|K|^{2(m+1)}2^m \prod_{1\leq i<j\leq m}\prod (\phi_i-\phi_j)$$

$$2^m|K|^{m+2} \prod_{1\leq i<j\leq m}\prod (\phi_i-\phi_j)$$

B.3 The technique of linear equivalence.

The identity expressed by equation (B.2.5) is extremely useful and should be thought of as a general technique rather than a "trick" to be applied only occasionally.

Let it be required to transform from K to $G = K^{-1}$ (both matrices mxm and elements funtionally independent) then

(B.3.1) $$KG = I.$$

Let a change δG in G produce a change δK in K then to first order

(B.3.2) $$(\delta K)G + K(\delta G) = 0$$

the derivative of K with respect to an element of G ($=K^{-1}$) is the coefficient of the corresponding element of δG. Thus,

(B.3.3) $$\left|\frac{\partial K}{\partial G}\right| = \left|\frac{\partial(\delta K)}{\partial(\delta G)}\right|$$

and since

(B.3.4) $$-G(\delta K)G = \delta G,$$

then

(B.3.5) $$\left|\frac{\partial(\delta K)}{\partial(\delta G)}\right| = |G \otimes G'| = |G|^{2m} = |K|^{-2m}$$

BIBLIOGRAPHY

An excellent bibliography is provided by T. W. Anderson in his Introduction to Multivariate Statistical Analysis (1958; John Wiley & Son, Inc., New York). Another source of references is provided by the International Journal of Abstracts (Published for "The International Statistical Institute" by Oliver and Boyd Ltd., Edinburgh). We give below references to publications immediately pertinent to these notes.

ANDERSON, T. W. (1955)
The integral of a symmetric unimodel function.
Proc. Amer. Math. Soc., 6, pp. 170-76.

ANDERSON, T. W. (1957)
Maximum likelihood estimates for a multivariate normal distribution when some observations are missing.
J.A.S.A., 52, pp. 200-03.

ANDERSON, T. W. (1958)
Introduction to Multivariate Statistical Analysis.
Wiley, New York.

ANDERSON, T. W. and BAHADUR, R. R. (1962)
Classification into two multivariate normal distributions with different covariance matrices.
Ann. Math. Stat., 33, pp. 420-31.

ANDERSON, T. W. and DAS GUPTA, S. (1963)
Some inequalities on characteristic roots of matrices.
Biometrika, 50, pp. 522-23.

ANDERSON, T. W. and DAS GUPTA, S. (1964)
Monotomicity of the power function of some tests of independence between two sets of variates.
Ann. Math. Stat., 35, pp. 206-08.

BANERJEE, D. P. (1958)
On the exact distribution of a test in multivariate analysis.
J.R.S.S., 20B, pp. 108-10.

BARTLETT, M. S. (1954)
A note on the multiplying factors for various χ^2 approximations.
J.R.S.S., 16B, pp. 296-98.

BENNETT, B. M. (1961)
Confidence limits for multivariate ratios.
J.R.S.S., 23B, pp. 108-12.

BHAT, B. R. (1962)
On the distribution of certain quadratic forms in normal variates.
J.R.S.S., 24B, pp. 148-51.

BISHOP, D. J. (1939)
On a comprehensive test of homogeneity of variances and covariances in multivariate problems.
Biometrika, 31, pp. 31-55.

BONNESEN, T. and FENCHEL, W. (1948)
Theorie der konvexen Körper.
Chelsea, New York.

BOX, G. E. P. (1949)
A general distribution theory for a class of likelihood criteria.
Biometrika, 36, pp. 317-46.

COCHRAN, W. G. (1934)
The distribution of quadratic forms in a normal system with applications to the analysis of variance.
Proc. Camb. Phil. Soc., 30, pp. 178-91.

CONSTANTINE, A. G. (1963)
Some noncentral distribution problems in multivariate analysis.
Ann. Math. Stat., 34, pp. 1270-85.

CONSTANTINE, A. G. and JAMES, A. T. (1958)
On the general cononical correlation distribution.
Ann. Math. Stat., 29, pp. 1146-66.

CURNOW, R. N. and DUNNETT, C. W. (1962)
The numerical evaluation of certain multivariate normal integrals.
Ann. Math. Stat., 33, pp. 571-79.

DAS GUPTA, S., ANDERSON, T. W. and MUDHOLKAR, G. S. (1964)
Monotonicity of the power function of some tests of the multivariate linear hypothesis.
Ann. Math. Stat., 35, pp. 200-05.

DAVID, F. N. (1938)
Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples.
Cambridge University Press.

DAVID, F. N. (1953)
A note on the evaluation of the multivariate normal integral.
Biometrika, 40, pp. 458-59.

DEEMER, W. L. and OLKIN, I. (1951)
The Jacobians of certain matrix transformations useful in multivariate analysis. Based on lectures of P. L. Hsu at the University of North Carolina.
Biometrika, 38, pp. 345-67.

FISHER, SIR RONALD A. (1962)
The simultaneous distribution of correlation coefficients.
Sankhyā, 24A, pp. 1-8.

GEISSER, S. (1963)
Multivariate analysis of variance for a special covariance case.
J.A.S.A., 58, pp. 660-69.

GEISSER, S. and GREENHOUSE, S. W. (1958)
An extension of Box's results on the use of the F-distribution in multivariate analysis.
Ann. Math. Stat., 29, pp. 885-91.

GHOSH, M. N. (1963)
Hotelling's generalized T^2 in the multivariate analysis of variance.
J.R.S.S., 25B, pp. 358-67.

GILLILAND, D. C. (1962)
Integral of the bivariate normal distribution over an offset circle.
J.A.S.A., 57, pp. 758-68.

GRAYBILL, F. A. (1961)
An Introduction to Linear Statistical Models, Volume I.
McGraw-Hill, New York.

GRAYBILL, F. A. and MARSAGLIA, G. (1957)
Idempotent matrices and quadratic forms in the general linear hypothesis.
Ann. Math. Stat., 28, pp. 678-86.

HALMOS, P. R. (1950)
Measure Theory.
D. van Nostrand, New York.

HANNAN, E. J. (1957)
The asymptotic processes of certain tests based on multiple correlations.
J.R.S.S., 19B, pp. 227-33.

HOTELLING, H. (1931)
The generalization of Student's ratio.
Ann. Math. Stat., 2, pp. 360-78.

HSU, P. L. (1938)
Notes on Hotelling's generalized T^2.
Ann. Math. Stat., 9, pp. 231-43.

HSU, P. L. (1945)
On the power functions for the E^2-test and T^2-test.
Ann. Math. Stat., 16, pp. 278-86.

IHM, P. (1959)
Numerical evaluation of certain multivarial normal integrals.
Sankhyā, 21, pp. 363-66.

IHM, P. (1961)
A further contribution to the numerical evaluation of certain multivariate normal integrals.
Sankhyā, 23A, pp. 205-06.

JACKSON, J. E. and BRADLEY, R. A. (1961)
Sequential χ^2-tests and T^2- tests
Ann. Math. Stat., 32, pp. 1063-77.

JAMES, A. T. (1955)
The non-central Wishart distribution.
Proc. Roy. Soc., A.229, pp. 364-66.

JAMES, A. T. (1960)
The distribution of the latent roots of the covariance matrix.
Ann. Math. Stat., 31, pp. 151-58.

JAMES, G. S. (1952)
Note on a theorem of Cochran.
Proc. Camb. Phil. Soc., 48, pp. 443-46.

JOHN, S. (1959)
On the evaluation of the probability integral of a multivariate normal distribution.
Sankhyā, 21, pp. 367-70.

KABE, D. G. (1963)
Multivariate linear hypothesis with linear restrictions.
J.R.S.S., 25B, pp. 348-51.

KABE, D. G. (1963)
Stepwise multivariate linear regression.
J.A.S.A., 58, pp. 770-73.

KENDALL, M. G. (1943 and 1946)
The Advanced Theory of Statistics I and II.
Charles Griffin and Co., London.

KHATRI, C. G. (1962)
Conditions for Wishartness and Independence of Second degree polynomials in a normal vector.
Ann. Math. Stat., 33, pp. 1002-07.

KSHIRSAGAR, A. M. (1961)
The non-central multivariate Beta distribution.
Ann. Math. Stat., 32, pp. 104-11.

LANCASTER, H. O. (1957)
Some properties of the bivariate normal distribution considered in the form of a contingency table.
Biometrika, 44, pp. 289-92.

LEVY, P. (1948)
The arithmetical character of the Wishart distribution.
Proc. Camb. Phil. Soc., 44, pp. 295-97.

MAHALANOBIS, P. C. (1936)
On the generalized distance in statistics.
Proc. Nat. Inst. Sci. India, 12, pp. 44-55.

MCFADDEN, J. A. (1960)
Two expansions for the quadravariate normal interval.
Biometrika, 47, pp. 325-33.

MIJARES, T. A. (1961)
The moments of elementary functions of roots of a matrix in multivariate analysis.
Ann. Math. Stat., 32, pp.1152-60.

OLKIN, I. (1952)
Notes on the Jacobians of certain matrix transformations useful in multivariate analysis.
Biometrika, 40, pp. 43-46.

OLKIN, I. and PRATT, J. W. (1958)
Unbiassed estimation of certain correlation coefficients.
Ann. Math. Stat., 29, pp. 201-211.

OLKIN, I. and RUBIN, H. (1962)
A characterization of the Wishart distribution.
Ann. Math. Stat., 33, pp. 1272-80.

PILLAI, K. C. S. and SAMSON, P., JR. (1959)
On Hotelling's generalized T^2.
Biometrika, 46, pp. 160-68.

RAO, C. R. (1959)
Some problems involving linear hypothesis in multivariate analysis.
Biometrika, 46, pp. 49-58.

RAO, C. R. (1960)
Multivariate analysis: An indispensible statistical item in applied research.
Sankhyā, 22, pp. 317-38.

RAO, K. S. (1951)
On the mutual independence of a set of Hotelling's T^2 derivable from a sample of size n from a k-variate normal population.
Bull. Inst. Internat. Stat., 33 (Part II), pp. 171-76.

RASCH, G. (1948)
A functional equation for Wishart's distribution.
Ann. Math. Stat., 19, pp. 262-66.

ROY, J. (1960)
Tests for independence and symmetry in multivariate normal populations.
Sankhyā, 22, pp. 267-78.

ROY, J. (1960)
Non-null distribution of the likelihood in analysis of dispersion.
Sankhyā, 22, pp. 289-94.

ROY, S. N. (1938)
Geometrical note on the use of rectangular coordinates in the theory of sampling distributions connected with a multivariate normal population.
Sankhyā, 3, pp. 273-84.

ROY, S. N. (1946)
On the power function of the different p-statistics for multivariate analysis of variance.
Proc. Indian. Sci. Cong.

ROY, S. N. (1953)
On a heuristic method of test construction and its use in multivariate analysis.
Ann. Math. Stat., 24, pp. 220-38.

ROY, S. N. (1957)
Some Aspects of Multivariate Analysis.
Wiley, New York.

ROY, S. N. and BARGMANN, R. E. (1958)
Test of multiple independence and the associated confidence bounds.
Ann. Math. Stat., 29, pp. 491-503.

ROY, S. N. and GNANADESIKAN, R. (1957)
Further contributions to multivariate confidence bounds.
Biometrika, 44, pp. 399-410.

ROY, S. N. and GNANADESIKAN, R. (1958)
A note on 'Further contributions to multivariate confidence bounds'.
Biometrika, 45, p. 581.

ROY, S. N. and ROY, J. (1959)
A note on a class of problems in normal multivariate analysis of variance.
Ann. Math. Stat., 30, pp. 577-80.

SONDHI, M. M. (1961)
A note on the quadravariate normal integral.
Biometrika, 48, pp. 201-03.

STEIN, C. M. (1962)
Confidence sets for the mean of a multivariate normal distribution.
J.R.S.S., 24B, pp. 265-96.

STUART, A. (1958)
Equally correlated variates and the multinormal integral.
J.R.S.S., 20B, pp. 373-78.

TANG, P. C. (1938)
The power function of the analysis of variance tests with tables and illustrations of their use.
Stat. Res. Mem., 2, pp. 126-57.

TUKEY, J. W. and WILKS, S. S. (1946)
Approximation of the distribution of the product of Beta variables by a single Beta variable.
Ann. Math. Stat., 17, pp. 318-24.

VAN DER VAART, H. R. (1961)
On certain characteristics of the distribution of latent roots of a symmetric matrix under general conditions.
Ann. Math. Stat., 32, pp. 864-73.

WALD, A. and BROOKNER, R. J. (1941)
On the distribution of Wilks' statistic for testing the independence of several groups of variates.
Ann. Math. Stat., 12, pp. 137-52.

WARNER, S. L. (1963)
Multivariate regression of dummy variates under normal assumptions.
J.A.S.A., 58, pp. 1054-63.

WEISS, L. (1958)
A test of fit for multivariate distributions.
Ann. Math. Stat., 29, pp. 595-99.

WIJSMAN, R. A. (1959)
Applications of a certain representation of a Wishart matrix.
Ann. Math. Stat., 30, pp. 597-600.

WILKS, S. S. (1946)
Sample criteria for testing equality of means, equality of variances and equality of covariances in a normal multivariate population.
Ann. Math. Stat., 17, pp. 257-81.

WISHART, JOHN (1928)
The generalized product moment distribution in samples from a multivariate population.
Biometrika, 20A, pp. 32-52.

WISHART, JOHN and BARTLETT, M. S. (1933)
The generalized product moment distribution.
Proc. Camb. Phil. Soc., 29, pp. 260-70.

WISHART, JOHN (1948)
Proof of the distribution law of the second order moment statistics.
Biometrika, 39, pp. 55-57.

CORRIGENDA

Some trivial errors, typically the ommission of "$\frac{1}{2}$" or a bracket, have been corrected in this volume. It is hoped that not too many errors of this type now exist.

The following corrections should be noted:

page 18. bottom left block of table.

for $\frac{n}{n-1}\underline{\bar{x}}'$...etc... read $n(n-1)\underline{\bar{x}}'$...etc...

page 18. bottom right block of table.

for $\dfrac{n\bar{x}^2}{\ldots\text{etc}\ldots}$ read $\dfrac{n\bar{x}^2}{\ldots\text{etc}\ldots.}$

page 45. line 6

for $\frac{1}{2}\nu(\nu+1)$ read $\frac{1}{2}\nu(2m-\nu+1)$

page 56. equation (3.4.20)

for $J = 2^m t_{11}^m t_{22}^{m-1} \ldots t_{m-1:m-1}^1 t_{mm}^0$

read $J = 2^m t_{11}^m t_{22}^{m-1} \ldots t_{m-1:m-1}^2 t_{mm}$

page 68. line 11 : denominator of f(T)

for $2^{\frac{1}{2}m(\nu-m)}$ read $2^{\frac{1}{2}m(\nu-2)}$

page 69. line 1 : denominator of f(R)

for $2^{\frac{1}{2}m(\nu-m)}$ read $2^{\frac{1}{2}m(\nu-2)}$

page 69. line 5 from bottom : denominator of $f(r_{ii})$

for $2^{\frac{1}{2}(\nu-i)-1}$ read $2^{\frac{1}{2}(\nu-i-1)}$

page 72. line 6

delete "$\frac{1}{2}$" in $\frac{1}{2}\,\mathrm{tr}W\frown$...etc...

CORRIGENDA (Cont'd)

page 75, line 5 of Theorem 33

for $\frac{|V|^h}{2^{nm}}$ ...etc... read $|V|^h 2^{hm}$...etc...

page 128. line 5 of Theorem 46

for $\hat{B}_\Omega \frown N_{m\times\ell}(B; V \otimes \Xi'(\Xi\Xi')^{-1}\Xi)$

read $\hat{B}_\Omega \frown N_{m\times\ell}(B; V \otimes (\Xi\Xi')^{-1})$

page 179. line 2 of Exercise 3

for $\frac{1}{(2\pi)\sqrt{1-\rho^2}}$ read $\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$

for exp $-\frac{1}{2}$...etc... read exp $-\frac{1}{2(1-\rho^2)}$...etc...

page 182. line 7

for "on $\underline{x}_{(2)}$ is replaced by $K\underline{x}_2$..."

read "on $\underline{x}_{(2)}$, that is when $\underline{x}_{(2)}$ is replaced by $K\underline{x}_{(2)}$..."

page 184. line 3

delete the summation signs

In the bibliography insert the references:

PITMAN, E. S. G. (1937)
The "closest" estimates of statistical parameters.
Proc. Camb. Phil. Soc., 33, p. 212.

SAW, J. G. (1964)
Some notes on variance-ratio tests of the general linear hypothesis.
Biometrika, 51, pp. 508-11.